数学建模与数学实验

主　编　贾　敬　桂占吉
　　　　蔡吉花
副主编　毋丽华　刘海林
主　审　何文章

哈尔滨工程大学出版社

数学建模与数学实验

主　编　贾　敬　桂占吉
　　　　蔡吉花
副主编　母丽华　刘海林
主　审　何文章

哈尔滨工程大学出版社

内 容 提 要

本书共分五章，主要包括微分方程模型、离散模型、数值分析模型、数学规划及数理统计模型。每章内容一般分为三部分：首先介绍模型实例，然后介绍建模所需的数学方法，最后介绍数学实验。每章均配有大量的数学建模实例，并配备了习题。

本书着重讲解数学建模的基本思想和方法，深入浅出、通俗易懂。以数学建模为主体，以计算机操作为手段，以数学软件为基础，把数学建模与数学实验紧密结合起来。该书可作为各类工科院校本专科数学改革教材，也可作为各行业工程技术人员的自学参考书。

数学建模与数学实验

SHU XUE JIAN MO YU SHU XUE SHI YAN

主　　编　贾　敬　桂占吉

　　　　　蔡吉花

副 主 编　毋丽华　刘海林

主　　审　何文章

责任编辑　张　奎

*

哈尔滨工程大学出版社出版发行

新　华　书　店　经　销

黑龙江矿业学院印刷厂印刷

*

开本　787×1092　1/16　印张 12.375　字数 271 千字

1998 年 7 月第 1 版　1998 年 7 月第 1 次印刷

印数：1～1000 册

ISBN 7-81007-882-8

O・62　定价：18.50 元

前　言

数学是工科大学生重要的基础理论课，它作为工程教育的一个重要内容，其目的在于培养工程技术人才所必备的基本素质。随着我国市场经济的不断发展，社会用人单位越来越感到我国的工科大学毕业生，理论联系实际的能力较低，反映在应用方面与发达国家的差距更大，针对这种情况，教育部曾委托全国工科数学课程指导委员会多次召开会议研究讨论工科数学教学改革问题，并于 1995 年 5 月下发了"关于工科数学系列课教学改革的建议"，按照这一建议的精神，结合当前国内外教学改革的发展趋势，我院制定了"工科数学教学内容与体系改革方案"。在近几年的教学改革试点中，我们开设了"数学建模与数学实验"课，根据教学中反馈出来的意见，我们又做了修改和补充，形成了现在这本教材。本教材作为这项改革的内容之一，其特色是突出数学建模的思想与方法，以数学建模内容为主体，以计算机操作为手段，以数学软件为基础，把数学建模与数学实验紧密结合起来。

本书由黑龙江矿业学院贾敬、桂占吉、蔡吉花任主编，黑龙江矿业学院母丽华、广东工业大学刘海林任副主编、佳木斯大学工学院鲁立刚也参加了本书的编写工作。

全书由黑龙江矿业学院何文章主审。

本书可作为工科院校本、专科数学改革教材，也可以作为参加数学建模竞赛培训的基础教材，还可供工程技术人员参考。

在本书的编写过程中得到了哈尔滨工程大学出版社及黑龙江矿业学院有关部门的大力支持和具体指导，得到了门艳春教授的关心和支持，在此表示衷心感谢。

由于编写时间仓促，水平有限，书中定有不当与错误之处，请各位同行与读者批评指正。

8

编　者

1998 年 7 月

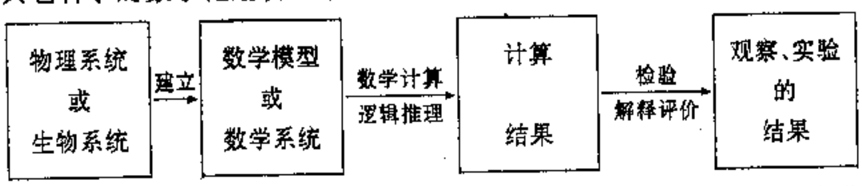
目　录

引言　数学模型概述

一、科学技术的数学化

现实世界中任何一种物质系统及其运动都有其质的规定性，又有其量的规定性，是质与量的统一. 由于数学所研究的量及其关系普遍存在于各种物质系统及其运动形式之中，因此一切科学技术原则上都可以用数学来解决相关的问题，其应用广度是空前的. 而用数学方法解决实际问题，首先都要通过抽象与简化，用数学的语言和方法，建立一个近似描述这个问题的数学模型. 数学模型是运用数学的语言和工具，对部分现实世界的信息（现象、数据、……）加以翻译、归纳的产物，它源于现实，又高于现实. 科学的发展离不开数学，数学模型在其中又起着重要作用. 科学就是通过对模型的研究来阐明真实世界的客观规律.

在近代，科学的数学化过程加快了，科学家越来越认识到数学在研究自然界中的重大作用，自牛顿开始，由于微积分的发明，力学的数学化便逐步臻于完善. 物理学的数学化过程走得更为深远. 当代最新的物理学许多重大成果都是数学工具的帮助下完成的. 它们的理论表述也往往采取了数学的形式，例如爱因斯坦的相对论力学就专门运用了黎曼在 19 世纪中叶创立的黎曼几何. 当代科学数学化的一个显著标志是数学向各门科学领域广泛渗透. 科学的数学化首先是在力学、物理学、化学、天文学、地质学等基础学科中进行的. 其次是生物科学的数学化. 生物科学是研究生命现象及其过程的. 在 19 世纪 80 年代中期，恩格斯根据当时的科学状况认为，数学在生物学中的应用等于零，到了 20 世纪，生物学向理论科学的转变过程中，使得它对数学的要求更迫切了，人们开始运用数学方法来研究生理现象，神经活动、生态系统以及遗传规律，并产生了数学生物学这样的学科，物理学与生物学以及其它科学的数学化途径基本上是一致的，我们可以用下图表示.



首先是从现实世界中的某一系统出发，这是科学研究的对象，它是一个十分复杂的系统，包含着许多变量，当我们选择了能在实验中被控制或可观察的参数，并用数学语言描述它们之间的关系时，一个数学模型便建立了，可见科学的数学化即意味着数学模型化.

社会科学最初是经济学的数学化，20 世纪 20 年代后期，西方经济学中产生了经济计量学这一学派. 这一学派在研究经济现象的数量关系时，采用了下列步骤：第一步建立经济数学模型，它是由表示经济现象的数量关系的方程组体系刻划的；第二步运用概率论和数理统计等数学方法估计参数；第三步把估算的参数值与观察值的统计资料相对照，进行理论验证；第四步根据已经估算出的参数值的方程式，求出预期变量的数值，即所谓预测

未来.目前,数学已广泛应用于社会科学的其它领域,如法学、人口学、史学等.尤其是计算机的发展,促进了科学技术数学化的进程和数学建模的发展,大量的计算是数学建模中不可缺少的,过去在高性能计算机尚未产生之前,正是由于缺乏这一技术手段而一定程度上限制了数学建模方法的应用和发展.而计算机的出现解决了数学建模中的大量而又复杂的计算问题.从而加速了科学数学化,使数学建模得到了飞速发展.

二、数学模型方法

数学模型方法,就是通过建立和研究客观对象的数学模型来揭示对象的本质特征和变化规律的一种方法.所谓数学模型是指对研究对象(现实原型)的本质特征和关系的数学表达.它是针对所要研究的具体事物的特征或数量相依关系,采用形式化的数学语言,概括地、近似地表达出来的一种数学结构,它或者能解释特定现象的现实状态,或者能预测对象的未来状况,或者能提供处理对象的最优决策或控制.

建立数学模型的方法一般要经过以下几个步骤:

1. 数学模型的建立

要了解所研究对象的实际背景,明确预期要达到的目标,根据研究对象的特点,确定刻画该对象系统的状态、特征和变化规律的若干基本量.这就要求我们查阅大量的资料,请教专家,力求掌握研究对象的各种信息,弄清实际对象的特征.

一个实际问题往往是很复杂的,影响它的因素比较多,如果想把全部影响因素都反映到建模中来,这样的数学模型就很难建立甚至是不可能建立的,即使建立了模型,也未必能达到预期的目标.因此,要根据研究对象的特征和建模目的,对问题进行必要的简化,这是建模的关键一步,不同的简化和假设会得到不同的模型.

考察一个质量为 m 的物体受到一个力 F 作用时的运动状况.如果物质运动速度不大,可以假设空气阻力可以忽略不计,根据牛顿第二定律,模型为

$$F = m\frac{d^2x}{dt^2}$$

这个式子表达了力 F 与位移 x 之间的规律.

若速度较大,假设空气阻力可以忽略不计就不合理了.必须考虑空气摩擦力这个因素,由实验得知,粘性阻尼的摩擦力与速度的平方成正比,于是得到一个该系统比较精确的数学模型.

$$F = m\frac{dv}{dt} + kv^2$$

因此,要针对所要解决的特定问题,分析系统中主要矛盾,摈弃可以忽略的因素,突出主要的因素和主要关系.

根据所做的假设,利用适当的数学工具,建立各个量之间的等式或不等式,画出图型列出表格或确定其它数学结构,建立模型.

2. 数学模型的求解与检验

模型的求解包括解方程、画图形、证明定理以及逻辑运算等,有时会用到传统的和近代的数学方法.

由于数学模型是进行一系列简化抽象的结果,但这种简化是否合理,其解是否能在误

差允许的范围内反映客观实际，这还是不得而知的. 因此求解和检验也是非常重要的一步，求解与检验的方法是:(1)对于比较简单的数学模型可以直接求解，安排相应的观察实验加以检验并根据检验结果加以修正.(2)运用计算机进行反复多次的“数学实验”. 过去有许多科学技术模型由于比较复杂，计算次数太多. 在计算机出现以前无法对其进行求解，对模型也无从验算和评价，随着计算机的出现和发展，提高了计算速度和求解方程的能力，为人们检验、校正、评价所建立的数学模型提供了有力的实验手段.(3)从理论上进行分析和探讨，对数学模型加以辅助性的检验. 因为有些模型如核战争模型是不可能接受实际的检验，对模型进行一些理论上的探讨也是很有实际意义的.

3. 对数学解的解释和评价

对数学解进行分析与评价，形成对实际问题的判断和预见，这是运用数学模型方法的最后一个步骤.

科学发展史上，根据数学模型求解的结果，做出科学的解释，提出重要的科学预言的事例很多. 英国物理学家狄拉克于 1928 年建立了相对论量子力学，给出了描写单个电子行为的电子波动方程——狄拉克方程，在解狄拉克方程时发现了电子负能态的存在，并且预言了正电子的存在，后来确实得到了证实. 当然对数学模型求解的结果要能做出正确的解释，形成对实际问题的判断和预见是很不容易的，这要求模型研究者具有广博的知识，丰富的想象力，深刻的洞察力，要勇于创新.

三、数学模型的分类

模型的分类比较复杂，分类方法很难有统一的标准，也没有严格界限，这里列举几种分类方法.

按照变量的情况，可分为确定性模型和随机性模型. 对于服从确定的因果联系，连续变化的自然过程，可以运用经典的数学方法，用各种方程式来描述，这类模型称为确定性模型. 自然界中还存在着另一类普遍现象，在这类现象中，就个别来看是无规律的，即事物的发展变化没有确定的因果性，有几种可能的不同结果，究竟出现哪一种结果是偶然的、随机的. 但是，当这类随机现象由大量成员组成或出现大量次数时，又可看出遵循着一种非偶然的规律性. 在许多系统中，由于受到一些复杂而尚未完全搞清楚的因素影响，使得系统在有确定的输入时，得到的输出是不确定的，这样的系统称为随机系统，这类数学模型为随机数学模型.

按照变量的情况，还可分为线性数学模型与非线性数学模型.

如果系统的输入与输出是呈线性关系，也就是说满足均匀性. 即满足若 $x_1(t)$，$x_2(t)$ 是任意两个输入;而 $y_1(t)$，$y_2(t)$为对应的输出.

$$x_1(t) \to y_1(t), x_2(t) \to y_2(t)$$

那么有

$$\alpha x_1(t) + \beta x_2(t) \to \alpha y_1(t) + \beta y_2(t)$$

我们称这种系统为线性系统，这种数学模型为线性数学模型，反之不满足上述关系式的，称之为非线性系统或非线性数学模型.

按变量情况还可分为连续模型与离散模型，单变量模型和多变量模型.

如果系统输入和输出是连续时间的函数，则称为连续时间系统或连续模型. 另一种系

统，它的输入与输出只在离散的瞬间取值，这种系统和模型称为离散时间系统或称为离散模型，通常用差分方程来表示.

按照时间变化对模型的影响，可分为定常数学模型和时变数学模型，若系统在初始条件给定的情况下，输出的状态取决于输入的状态而与输入的时刻无关，具有这种特性的系统称为定常系统，对应的数学模型叫做定常数学模型. 反之，若参数为时间 t 的函数时，系统为时变系统，相应的模型称为时变数学模型.

按照研究方法和对象的数学特征分类，有初等模型、优化模型、逻辑模型、稳定性模型、模糊数学模型、突变模型等.

按照研究对象的实际领域，有人口模型、交通模型、生态模型、经济模型、社会模型、体育模型等.

模型的分类问题在本门课程中并没有什么重要意义，我们应该把注意力集中在建立模型，运用模型解决实际问题上来，从而达到增强数学的应用能力和加强数学素质的培养的目的.

第一章　微分方程模型

在研究实际问题时，经常无法找到该系统的有关变量之间的直接关系.但是却容易找到这些变量和它们的变化率之间的关系式，利用这些微分关系式，我们可以建立相应的微分方程模型.这种模型揭示出系统的变量在局部或瞬时所遵循的规律.因此在解决实际问题时，微分方程模型得到了广泛应用.

第一节　放射性元素衰变模型

一、马王堆一号墓的年代

湖南省长沙市马王堆一号墓于1972年8月出土，估算出该墓的大致年代是考古工作者的重大课题.

美国物理学家Rutherford和他的同事们证明了某些放射性元素的原子是不稳的，在一个给定的时期内，一定比例的原子会自动衰变，形成新元素的原子.Rutherford还证明了，放射性物质在任意时刻的衰变速度都与该物质现存的数量成正比.

对有机物(动、植物)遗体的年代测定方法是放射性碳14测定法.它的根据是：由于大气层受到宇宙射线的连续照射，使之产生中子，中子与空气中的氮结合生成具有放射性的碳14(C^{14}).有机体存活时，它们通过新陈代谢与外界进行物质交换，使体内的C^{14}与空气中的C^{14}有相同的百分含量.一旦有机体死亡，新陈代谢终止，因而体内C^{14}由于不断地衰变而不断减少.这种测定法就是根据衰变减少量的变化情况来判定有机体的死亡时间.

设在时间t(年)碳14的存量为y，由上述原理可得微分方程

$$\frac{\mathrm{d}y}{\mathrm{d}t} = -\lambda y \tag{1.1}$$

其中$\lambda>0$为衰变常数，式中负号表示C^{14}的存量y是递减的.对于物质衰变速度的一种度量就是物质的半衰期，它定义为一定数量的放射原子衰变到一半时所需要的时间.设有机体死亡时间为$t_0=0$，C^{14}含量为y_0，则初值问题

$$\frac{\mathrm{d}y}{\mathrm{d}t} = -\lambda y, y(0) = y_0 \tag{1.2}$$

的解是

$$y(t) = y_0 e^{-\lambda t} \tag{1.3}$$

由(1.3)式可得

$$t = \frac{1}{\lambda}\ln\frac{y_0}{y} \tag{1.4}$$

由元素衰变的半衰期$y(T_{1/2})=\frac{y_0}{2}$决定衰变常数为

$$\lambda = \frac{\ln 2}{T_{½}} \tag{1.5}$$

由于

$$y'(0) = -\lambda x(0) = -\lambda y_0 \tag{1.6}$$

所以

$$\frac{y'(0)}{y'(t)} = \frac{y_0}{y(t)} \tag{1.7}$$

将(1.5),(1.7)式代入(1.4)式,得

$$t = \frac{T_{1/2}}{\ln 2}\ln\frac{y'(0)}{y'(t)} \tag{1.8}$$

其中 $y'(0)$表示有机体死亡时C^{14}的变化率,由于地球周围的C^{14}的百分含量可以认为是基本不变的,因而认为现代有机体中C^{14}的衰变速度与古代有机体中C^{14}的衰变速度相同,所以可以用现代同类有机体死亡时C^{14}的变化率代替 $y'(0)$. 我们可以测得出土木炭标本的C^{14}的 $y'(t)$为 29.78 次/分,$y'(0)$为 38.37 次/分,C^{14}的半衰期为 5 730 年,将这些数据代入(1.8)式,得

$$t = \frac{5730}{\ln 2}\ln\frac{38.37}{29.78} \approx 2095(\text{年})$$

这样就估算出马王堆一号墓的年代大约是在 2 000 年前的.

二、赝品的鉴定

Jan Vermeer 是 17 世纪荷兰著名画家,他的传世作品仅有 37 件. 他生前卖出的画很少,以致在相当长的历史时期内他的画都没有引起人们的重视. 直到 19 世纪中叶,人们才发现了这位杰出的画家. 由于他的画严谨细腻,精美绝伦,使他的画价剧增,成为当时文物市上售价最高的作品之一. 第二次世界大战中,比利时解放后,德国战场安全部在追捕纳粹同党时,引发了 Van Meegeren 伪造 Jan Vermeer 的名画案.

德国战场安全部在调查一家曾把大批艺术品出卖给德国的公司中,发现一位银行家. 这位银行家在拍卖 Jan Mermeer 的名画中作过中间人. 他还承认,他曾是德国三流画家 H. A. Van Meegeren 的代理人. 1949 年 5 月 29 日,Meegeren 以通敌罪被逮捕. 然而,在 1945 年 7 月 12 日,Meegeren 在监狱中说,他从未拍卖过 Vermeer 的名画.《妓女》以及非常著名和精美的《Emmaus 的信徒们》,还有其它四幅被认为是 Vermeer 的作品,都是他伪造的. 许多人认为 Van Meegeren 是为了逃脱通敌罪而说谎. 为了证实他所说的一切,他在监狱中开始伪造 Vermeer 的画《耶稣在学者们中间》,从而向持怀疑态度的人表明自己是一个高明的伪造者. 当他的工作几乎要完成时,他获悉他的通敌罪可能被伪造罪所取代. 于是,他拒绝完成他的画及使之老化,他希望调查人员不会发现使他的伪造品老化的秘密. 为了解决这个问题,一个由著名化学家、物理学家和艺术家组成的国际调查小组受命调查此事. 调查小组对油画作了 X-射线检查以确定是否有其它油画在它们下面. 此外,他们分析了绘画所用的颜料,并检验某些年代迹象. 尽管 Van Meegeren 伪造的十分逼真,专家调查小组还是在油画中发现了现代颜料钴蓝的痕迹. 在有些画中还发现了 19 世纪才发明的苯酚甲醛,因此伪造罪成立. Van Meegeren 被判处有期徒刑一年. 在服刑期间,他因心脏病发作于 1947 年 11 月 30 日死亡.

尽管专家调查小组提供了某些证据，许多人还是不相信著名的《Emmaus 的信徒们》是 Van Meegeren 伪造的．甚至著名的艺术史学家 A. Bredius 也认为《Emmaus 的信徒们》是真正的 Vermeer 的作品．这幅画已被 Rembradt 协会用 170 000 美元买下．1967 年卡内基-梅隆大学的科学家们给出了一个完全科学的、确定性的论据，证明了《Emmaus 的信徒们》确实是一件伪造品．

由初等化学知识可以知道，地壳中的所有岩石几乎都含有少量的铀．岩石中的铀衰变成一种其它的元素，而这种元素又衰变成另一种元素，如此衰变下去，形成一个元素系列直到铅就不再衰变了．铀不断地补充序列中的后续元素．所以，这些后面的元素被前面的元素替代的速度与他们衰变的速度同样快．

画家使用白铅做为颜料已有 2 000 多年的历史．白铅属于铀系，其主要成分是无放射性的白铅 206，也含有微量的放射性元素镭 226 和铅 210．白铅是由铅金属产生的，而铅金属是经过熔炼从铅矿石中提取出来的．在这个过程中，矿石中的铅 210 随铅金属被提取出来．不过 90%～95% 的镭以及它的派生物都随着炉渣中的废物被排出了．所以大多数铅 210 的提供物被排掉了，而铅 210 开始迅速地衰变，其半衰期为 22 年．这个衰变过程一直持续到白铅中的铅 210 再次与现存的少量镭达到放射平衡，即铅 210 的衰变数恰好被镭的衰变所平衡．

下面来计算从油画颜料样本中得到的铅 210 的含量，设 $y(t)$ 为 t 时每克白铅中铅 210 的数量，y_0 为油画刚完成的 t_0 时刻每克白铅中存在的铅 210 的数量，而 $r(t)$ 为 t 时刻每分钟每克白铅中镭 226 的衰变数．λ 是铅 210 的衰变数，则 $y(t)$ 是如下初值问题的解．

$$\frac{\mathrm{d}y}{\mathrm{d}t} = -\lambda y + r(t), y(t_0) = y_0 \tag{1.9}$$

由于我们感兴趣的只是 300 年左右的事情，而镭 226 的半衰期是 1 600 年，故 $r(t)$ 是一个常数，即 $r(t)=r$，解方程(1.9)得

$$y(t) = \frac{r}{\lambda}(1 - \mathrm{e}^{-\lambda(t-t_0)}) + y_0\mathrm{e}^{-\lambda(t-t_0)} \tag{1.10}$$

通过下面方法，我们可以为赝品的鉴定提供科学论据．

若画是真品，我们选 $t-t_0\approx300$，则

$$\lambda y_0 = \lambda y(t)\mathrm{e}^{\lambda(t-t_0)} - r[\mathrm{e}^{\lambda(t-t_0)} - 1]. \tag{1.11}$$

表 1-1　被怀疑为赝品的油画

画　　　名	钋 210 的衰变	镭 226 的衰变
Emmaus 的信徒们	8.5	0.8
洗　足	12.6	0.26
读乐谱的妇人	10.3	0.3
弹曼陀林的妇人	8.2	0.17
做花边的人	1.5	1.4
欢笑的女孩	5.2	6

$$\lambda y_0 \approx \lambda y(t)\mathrm{e}^{300\lambda} - r(\mathrm{e}^{300\lambda} - 1) \tag{1.12}$$

又由于铅 210 的半衰期为 22，由(1.5)式得 $\lambda=\ln2/22$，因此

$$e^{300\lambda} = e^{\frac{300}{22}\ln 2} = 2^{\frac{150}{11}} \tag{1.13}$$

由于钋 210 的衰变率等于铅 210 的衰变率，而且钋 210 的衰变率较易测量．人们测定了《Emmaus 的信徒们》及其它几幅被认为是伪造品的钋 210 和镭 226 的衰变率如表 1-1．我们由(1.12)计算《Emmaus 的信徒们》中白铅的 λy_0，得

$$\lambda y_0 = 2^{\frac{150}{11}} \cdot 8.5 + 0.8 \cdot (2^{\frac{150}{11}} - 1) = 98\ 050 \text{ 个}$$

这个值大得令人无法接受，因此，这幅《Emmaus 的信徒们》一定是伪造的．用类似的分析可以证明《洗足》、《读乐谱的妇人》、《弹曼陀林的妇人》都是伪造 Vermeer 的．另一方面专家们也证实了《做花边的人》和《欢笑的女孩》不可能是近期伪造 Vermeer 的画．因为这两幅画的钋 210 与镭 226 几乎接近放射平衡，而且在 19 世纪或 20 世纪的绘画样品中都没有观察到这样的平衡．

第二节　恶性肿瘤增长模型

恶性肿瘤是一种癌症．人类为了攻克癌症，医学、遗传学、生物学、生理学、生物物理、生物化学等各科的专家们进行各种联合，以便研究它的发生成长规律及防治方法．随着医疗卫生、科学技术的不断创新，蓬勃发展，癌的防治工作也发生了令人瞩目的进展．癌的发展包括两个阶段：转形变异和增大．在转形变异阶段，正常的细胞会变成一个有能力生成肿瘤的异常细胞．在增长阶段异常细胞的分裂速率比正常细胞高得多，进行分裂倍增，从而导致肿瘤的增长发展．对恶性肿瘤，异常细胞的增长常常是侵略性的，它们甚至可以杀死宿主．本节介绍了四个恶性肿瘤增长模型．

一、指数模型

设 $N(t)$ 为时刻 t 的肿瘤细胞数目，设肿瘤细胞的相对增长率为 λ，如果 $N(t)$ 相对 t 的变化率是与 $N(t)$ 的瞬时值成正比，则

$$\frac{dN}{dt} = \lambda N$$

由于肿瘤细胞数目超过 10^{11} 时，才是临床可观察的，所以 $N(0)=10^{11}$．$N(t)$ 满足初值问题

$$\frac{dN}{dt} = \lambda N, N(0) = 10^{11}$$

的解是

$$N(t) = 10^{11}e^{\lambda t}$$

设肿瘤细胞增加一倍所需时间为 T，即 $N(t+T)=2N(t)$，由此可得

$$\lambda = \ln 2/T$$

这就是肿瘤增长的指数模型．大量临床实践表明：在肿瘤生长初期，几乎每经过一定的时间，肿瘤细胞数目就增加一倍．在肿瘤生长后期，由于各种生理条件的限制，肿瘤细胞数目逐渐趋向于某个稳定值．所以指数模型适应于肿瘤生长初期，后期会出现较大偏差．

二、Logistic 模型

用荷兰生物数学家 Verhulst 提出的 Logistic 模型更为合适．记由于生理限制肿瘤细

胞数目的极限值为 N_m，关于 $N(t)$的微分方程可写作

$$\frac{dN}{dt}=\lambda N\frac{N_m-N}{N_m-N(0)} \qquad (1.14)$$

解为

$$N(t)=N(0)\left[\frac{N(0)}{N_m}+\left(1-\frac{N(0)}{N_m}\right)e^{-\frac{\lambda N_m t}{N_m-N(0)}}\right]^{-1}$$

下面我们研究 Logistic 模型的解曲线的凹凸性及拐点. 对(1.14)式求导，得

$$\frac{d^2N}{dt^2}=\lambda\frac{dN}{dt}\left(\frac{N_m-N}{N_m-N(0)}\right)+\lambda N\left(-\frac{1}{N_m-N(0)}\frac{dN}{dt}\right)=\lambda\frac{dN}{dt}\left(\frac{N_m-2N}{N_m-N(0)}\right).$$

显然拐点的横坐标为$\frac{N_m}{2}$，当 $N<\frac{N_m}{2}$时，二阶导数为正，曲线 $N(t)$是凹的；当$\frac{N_m}{2}<N<N_m$时，二阶导数为负，曲线 $N(t)$是凸的. 并且曲线的凹凸弧段基本是关于拐点对称的. 专家在实践中发现，肿瘤生长曲线一般不是关于 Logistic 模型的拐点对称. 即拐点横坐标未必取在$\frac{N_m}{2}$处. 因此 Logistic 模型经常不能与实际数据很好地吻合.

三、Gompertz 模型

$N(t)$、N_m 及 λ 表达的意义同 Logistic 模型，则 Gompertz 模型为

$$\frac{dN}{dt}=-\lambda N\ln\frac{N}{N_m} \qquad (1.15)$$

方程(1.15)的解为

$$N(t)=N(0)\exp\left[\ln\frac{N_m}{N(0)}(1-e^{-\lambda t})\right]$$

显然方程的解有以下性质：

(1)当 $t\to\infty$时，$N(t)\to N_m$.

(2)当 t 较小时，$e^{-\lambda t}\approx 1-\lambda t$，从而得 $N(t)\approx N(0)e^{(\ln\frac{N_m}{N(0)})\lambda t}$. 所以当 t 较少时，Gompertz 模型与指数模型近似.

(3)当 $N<\frac{N_m}{e}$时，$\frac{d^2N}{dt^2}>0$；当 $N>\frac{N_m}{e}$时，$\frac{d^2N}{dt^2}<0$；而 $N=\frac{N_m}{e}$是解曲线拐点的横坐标.

由于 Gompertz 模型的解曲线的拐点横坐标为$\frac{N_m}{e}$，凹凸弧段不是关于拐点对称，所以此模型能比较灵活地与实际数据吻合.

J. R. Usher 提出了一个更一般的模型：

$$\frac{dN}{dt}=\frac{\lambda N}{\alpha}\left[1-\left(\frac{N}{N_m}\right)^{\alpha}\right]$$

它的解为

$$N(t)=N(0)\left\{\left(\frac{N(0)}{N_m}\right)^{\alpha}+e^{-\lambda t}\left[1-\left(\frac{N(0)}{N}\right)^{\alpha}\right]\right\}^{-\frac{1}{\alpha}}$$

上述三个模型均为 Usher 模型的特殊形式：

当 $\alpha\to 0$ 时，$\frac{dN}{dt}\to-\lambda N\ln\frac{N}{N_m}$，即得 Gompertz 模型；

当 $\alpha\to 1$ 时，$\frac{dN}{dt}\to\lambda N\left(1-\frac{N}{N_m}\right)$，即得 Logistic 模型；

当 $\alpha\to1$ 且 $N_m\to\infty$ 时，$\frac{dN}{dt}\to\lambda N$，即得指数模型.

四、与时间有关的肿瘤生长模型

在指出增长模型 $\frac{dN}{dt}=\lambda N$ 时，把相对增长率 λ 换成 $\lambda e^{-\alpha t}$，这表示肿瘤细胞分裂增殖的平均时间逐渐增加，使得肿瘤增长速度减慢. 这是因为生殖细胞占的比例虽然没变，但它们逐渐老化了，分裂变慢，此时模型为

$$\frac{dN}{dt}=\lambda e^{-\alpha t}N \tag{1.16}$$

其中 λ,α 均为正数. $N(t)$ 和 $N(0)$ 的意义与上述其它模型一致，则方程的解为

$$N(t)=N(0)\exp\left\{\frac{\lambda}{\alpha}[1-\exp(-\alpha t)]\right\}$$

由上式可见，肿瘤细胞数目的增长速度越来越慢，当时间趋向无穷大时，极限为 $N(0)e^{\frac{\lambda}{\alpha}}$.

某些医学专家对与时间有关的肿瘤增长模型提出了如下解释.

方程(1.16)可以写成

$$\frac{dN}{dt}=\lambda(e^{-\alpha t}N)$$

与指数模型比较，将 N 改成 $e^{-\alpha t}N$，表示肿瘤的生殖细胞随时间逐渐减少，使得肿瘤增长变慢. 系数 λ 没有改变，说明分裂细胞增殖的平均时间没变. 对于这种现象，被解释为随着肿瘤的增大，向中心核供氧越来越困难，中心会出现坏死现象，使生殖细胞逐渐减少.

第三节　人体减肥模型

众所周知，肥胖人群患心脑血管疾病、糖尿病、肺心病的比例比正常人明显偏高. 随着人们生活水平的提高，肥胖症患者和超过标准体重的人越来越多. 在我国中小学生中，肥胖学生比例也有增高的趋势. 减肥问题引起了人们广泛关注. 目前各种减肥药物和减肥器械纷纷登场，但疗效并不令人满意，收效甚微. 我们在这里建立了减肥的两个数学模型，仅从数学上对减肥问题作些初步探讨.

一、线性模型

首先我们介绍一个实际例子，从这个实例的解决过程，概括出具有一般意义的线性模型.

例 1　某人的食量是每天 2 500 卡，其中 1 200 卡用于基本新陈代谢. 在健身训练中，每天他的每千克体重消耗 16 卡. 假设以脂肪形成贮藏的热量 100% 地有效，而 1 千克脂肪含热量 10 000 卡，求出这个人的体重是怎样随时间变化的.

在这个问题中并没有“导数”这个十分关键的词，但问题是求这个人的体重 w 关于时间的函数. 如果我们把 w 看成是时间 t 的连续函数，我们就找到一个含有 $\frac{dw}{dt}$ 的微分方程. 首先我们列出概念性陈述.

每天，重量的变化 = 输入 − 输出.

输入是指扣除了基本新陈代谢之外的净重量吸收.

输出是指进行了健身训练时的消耗.

由于我们的目的是考察导数，因此，上面的陈述可以合并为更好的概念陈述：

体重变化/每天=净吸收量/每天−消耗/每天.

每天的净吸收量=2 500 卡/D−1 200 卡/D=1 300 卡/D

每天的净输出=16w 卡/D

体重的变化/D=$\frac{dw}{dt}$千克/D

式中 D 表示每天，我们注意到，有些量的单位是卡，有些量的单位用千克. 题中给出了转化为 1 千克脂肪含热量 10 000 卡. 从而得到了微分方程

$$\frac{dw}{dt}=\frac{(2\,500-1\,200)-16w}{10\,000}$$

初始条件为一天的开始时他的体重：$t=0, w=w_0$，初值问题的解为

$$w=\frac{1\,300}{16}-\frac{1\,300-16w_0}{16}\exp(-16t/10\,000)$$

这样就回答了所提出的问题. 现在，我们再来考虑一个问题：这个人的体重会达到平衡吗？

显然当 $t\to\infty$ 时，$w\to\frac{1\,300}{16}$千克=81.25 千克.

建立模型时一般要给出一些假设，根据上述讨论，我们做出如下假设：

1. 假设人体的骨骼、内脏等变化不大，视为常数，人体重的增减仅仅是脂肪的增减.

2. 由于一个人生活习惯相对稳定，所以假定每天所吃食物产生的热量为常数 p 卡.

3. 每天用于基本的新陈代谢的热量为 q 卡.

4. 假设每天因运动消耗的能量与体重成正比，即 $mw(t)$. 其中，人的体重为 $w(t)$ 千克，人每天因运动每千克体重消耗 m 卡.

5. 以脂肪形式贮藏的热量 100%有效，1 千克脂肪含热量 n 卡.

6. 对于人体系统而言，能量守恒.

7. 初始时刻 $t=0$ 时，体重为 w_0 千克.

由上述假设，根据能量守恒原理在 Δt 时间内恒成立

$$[w(t+\Delta t)-w(t)]\cdot n=[p-q-mw(t)]\cdot\Delta t$$

记 $\frac{p-q}{n}=a, \frac{m}{n}=b$，则 $a\geqslant 0, b>0$ 且上式变为

$$\frac{w(t+\Delta t)-w(t)}{\Delta t}=a-bw(t) \tag{1.17}$$

在(1.17)式中令 $\Delta t\to 0$，得到如下减肥模型

$$\frac{dw}{dt}=a-bw, w(0)=w_0, \tag{1.18}$$

解为

$$w(t)=\frac{a}{b}-\frac{a-bw_0}{b}e^{-bt}$$

当 $t\to\infty$ 时，$w(t)\to\frac{a}{b}$. 所以这个人的体重达到平衡时的值为 $\frac{a}{b}$. 我们还可以直接从微分方程(1.18)求出达到平衡状态的平衡值. 因为在平衡状态下，w 不发生变化，所以 $\frac{\mathrm{d}w}{\mathrm{d}t}=0$. 这就非常直接地给出了平衡值 $\tilde{w}=\frac{a}{b}$ 千克. 由此可见，如果我们需要知道的仅仅是这个平衡值，不必解微分方程就可以得到它.

从模型结论可以看出，通过减小 a 或者增大 b 可达到减肥的目的. 减小 $a=\frac{p-q}{n}$，即适当控制饮食，减少高脂肪食物的摄入，可以控制体重的增加. 增大 $b=\frac{m}{n}$，即增加运动量，甚至每天都要进行健身训练可减轻体重. 在饮食规律一定的前题下，每天都进行健身训练，永远坚持下去就可使体重达到一个平衡值.

二、非线性模型

在上述讨论中，我们在第 4 条假设中，假设每天因运动消耗的能量与体重成正比. 经专家研究表明运动消耗能量并非都是这种简单的线性关系. 因此，我们还可以假设每天因运动消耗的能量与 $w^{\frac{3}{2}}$ 成正比，这样就得到如下模型：

$$\frac{\mathrm{d}w}{\mathrm{d}t}=a-bw^{\frac{3}{2}},w(0)=w_0.$$

这是一个非线性方程. 若我们只想知道这个人的体重是否会达到平衡，平衡值是多少？我们可以讨论方程

$$\frac{\mathrm{d}w}{\mathrm{d}t}=a-bw^{\frac{3}{2}}$$

令 $\frac{\mathrm{d}w}{\mathrm{d}t}=0$，可得平衡值为 $\tilde{w}=\left(\frac{a}{b}\right)^{\frac{2}{3}}$. 若我们还想知道人的体重随时间变化的全过程，就得讨论方程初值问题的解. 令 $w=u^2$，则非线性方程化为变量可分离的微分方程，解之得.

$$t=\frac{2}{36l}\ln\left|\frac{\sqrt{w}+l}{\sqrt{w}-l}\right|+f(w)$$

其中 $l=\sqrt[3]{\frac{a}{b}}$，.

$$f(w)=\frac{2}{36l}\left[\operatorname{arctg}\left(\frac{2\sqrt{w_0}}{\sqrt{3}\,l}-\frac{1}{\sqrt{3}}\right)-\operatorname{arctg}\left(\frac{2\sqrt{w}}{\sqrt{3}\,l}-\frac{1}{\sqrt{3}}\right)\right]-\frac{2}{36l}\ln\left|\frac{\sqrt{w_0}+l}{\sqrt{w_0}-l}\right|$$

很容易地得到估计式

$$|f(w)|\leqslant\frac{2}{36l}\left(\pi+\ln\left|\frac{\sqrt{w_0}+l}{\sqrt{w_0}-l}\right|\right)$$

于是，我们还可以得到 $\lim\limits_{t\to+\infty}w(t)=\left(\frac{a}{b}\right)^{\frac{2}{3}}$.

第四节　新产品销售模型

一种产品进入市场后，经营者自然要关心消费者以怎样的速度接受这种产品，以及广告在销售中的效果等销售规律. 本节将介绍三个销售模型来初步探讨销售规律.

一、新产品销售的 Logistic 模型

假定在 $t=0$ 时，一种新产品被介绍到一个需求量的上界为 K 的市场里. 设 $N(t)$ 表示 t 时刻购买了新产品的人数. 尽管 $N(t)$ 是以整数变化的，但我们仍把它近似地视为时间的连续函数. 我们假设：对于一位尚未购买新产品的潜在消费者来说，只有当一位产品购买者对他谈论了这种产品后，他才可能会购买它. 于是，在很短的时间间隔 Δt 内采纳这种产品的消费者数 ΔN 与在此之前已经购买了产品的消费者数 N 以及还不知道这种产品的潜在消费者 $K-N$ 成正比. 因此在比例系数为 λ 时，有

$$\Delta N=\lambda N(t)(K-N(t))\cdot\Delta t$$

或

$$\frac{\mathrm{d}N}{\mathrm{d}t}=\lambda N(t)(K-N(t)) \tag{1.19}$$

设 $t=0$ 时，有 N_0 个潜在消费者购买了这种产品. 我们可以得到满足初值问题

$$\frac{\mathrm{d}N}{\mathrm{d}t}=\lambda N(K-N),N(0)=N_0 \tag{1.20}$$

的解为

$$N(t)=\frac{Ke^{\lambda Kt}}{K-N_0+e^{\lambda Kt}}$$

此即为 Logistic 模型. 从它的函数图象可知，当销售量小于最大需求量一半时，销售速度是逐渐增加的；销售量恰好达到最大需求量一半时，该产品最为畅销，其后销售速度开始下降. 实际情况表明，产品销售情况有时与 Logistic 模型十分吻合.

二、Bass 模型

在 Logistic 模型中，我们假设对于一位尚未购买新产品的潜在消费者来说，只有当一位产品购买者对他谈论了这种产品后，他才可能购买它. 因此 Logistic 模型有时与实际不符 Bass 认为新产品进入市场后的扩散过程类似于传染病的扩散过程. 他以传染病模式为基础导出了一个微分方程模型，用以估计包括黑白电视机等多种耐用消费品的销售量. Bass 认为一个新型产品进入市场，其有关信息的传播一般有两个途径. 一部分人购买之后经过使用而对产品有所评价并传播开来，在这些早期产品购买者周围的人们也就得到了有关产品的信息. 再者是经营者或厂家通过大量的宣传工具，

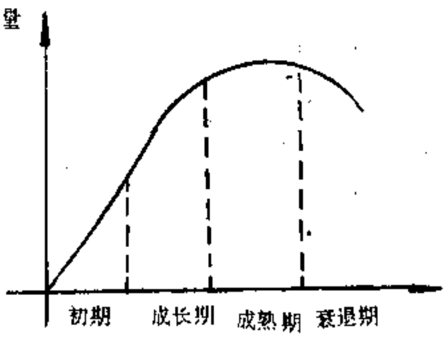


图 1-1　产品生命周期曲线

如收音机、电视、报纸等上面的广告宣传，或人们在商店亲眼看到商品等来自购买者以外的信息. 下面给出 Bass 模型的表示.

设 K 为潜在的消费者总体数，$N(t)$为时刻 t 购买了新产品的人数. 其中 $K \geqslant N(t)$. 一部分是来自消费者内部口头传播的有关产品信息，导致的购买者增量.

$$\lambda N(t)(K - N(t))\Delta t$$

其中 λ 是大于零的比例系数，如果不考虑另外一种传播途径，则可得到新产品销售的 Logistic 模型. 如果把来自消费者外部的有关产品的信息导致的购买者因素考虑进去，在 Δt 时，通过广告宣传和其他途径促使潜在消费者可能购买数与还不知道这种新产品的潜在消费者数成正比，即对某个正常数 μ

$$\mu(K - N(t))\Delta t$$

所以，在时间段 Δt 内，购买者增量为

$$\Delta N = \lambda N(t)(K - N(t))\Delta t + \mu(K - N(t))\Delta t.$$

从而
$$\frac{\mathrm{d}N}{\mathrm{d}t} = \lambda N(K - N) + \mu(K - N(t))$$

这就是 Bass 模型，如果 $N(0)=0$，则 $N(t)$满足初值问题的解为

$$N(t) = \frac{K\mu[\mathrm{e}^{(\mu+\lambda K)t} - 1]}{\lambda K + \mu \mathrm{e}^{(\mu+\lambda K)t}}$$

一种产品进入市场后，一般会经过一个销售量不断增加然后又逐渐下降的过程，称为产品的生命周期，简记为 PLC(Product Life Cycle). Bass 模型成功地区分了同一消费者总体中产品信息的两种不同的传播渠道. Bass 的研究还进一步将消费者总体分为两部分，一部分叫做革新者，另一部分叫做模仿者. 他认为在同一社会里，一些人对新型的商品、技术的采用总是比另一些人要早些，于是被称为革新者，而那些较晚采用的人就被称为模仿者. 人们接受了 Bass 提出的信息传播的两种途径及把消费者分为革新者和模仿者的概念. 但 Bass 模型得到的 PLC 曲线是单峰的，Bass 模型看不到他所分析的两类不同消费者在市场中表现出的不同消费行为，但恰恰是消费者的行为决定了新产品推广模式. 但对于新型耐消费品，我们观察到，有时其生命曲线在开始阶段有一个小的销售高峰，然后有一段时间的持平或销量下降，而后再次进入销售上升阶段，达到高峰. 从而出现双峰的 PLC. 这类问题将由下面的模型解决.

三、Steffens-Murthy 模型

P. R. Steffens 和 D. N. P. Murthy 针对 Bass 模型存在的缺点. 以革新者和模仿者在市场中表现出的不同消费行为为依据建立了新的产品销售模型. Steffens-Murthy 模型的 PLC 曲线不仅能出现单峰，同时也适用于出现双峰的情形. 首先我们先给出模型的假设：

1. 有关新产品信息可分为两类：搜寻型信息和体验型信息. 搜寻型信息是指人们从广告、光顾商店以及从产品的使用说明书了解和获得的信息. 体验型信息是指消费者购买产品使用产品得到的对产品的认识信息.

2. 消费者总体按照他们在决定采用新产品之前所需要的信息分成两部分. 第一部分是消费者有了产品的搜寻型信息就决定购买该产品，称为革新者；第二部分是消费者一定

要得到产品体验型信息之后才购买该产品，称为模仿者.

3. 消费者总体人数为 K，δ_1 和 δ_2 分别为革新者和模仿者的比例系数，则

$$K_1 = \delta_1 K = \text{革新者总数}, K_2 = \delta_2 K = \text{模仿者总数},$$

且 $\delta_1+\delta_2\leqslant 1$.

4. 假设无论是革新者还是模仿者对搜寻型信息的传播速度是一致的.

5. 假设模仿者一旦得到了产品使用的体验型信息后就决定购买.

下面建立搜寻型信息传播的模型.

设 $F_1(t)$和 $F(t)$分别表示在 t 时刻已得到搜寻型信息的革新者人数和总人数. 由假设 4，有

$$F_1(t) = \delta_1 F(t) \tag{1.21}$$

考虑到搜寻型信息的传播仍有两个途径，即来自消费者外部的广告传播及来自消费者内部的口头传播. 仿照 Bass 模型的做法，给出搜寻型信息的传播公式

$$\frac{\mathrm{d}F(t)}{\mathrm{d}t} = a_1 F(t)(K - F(t)) + a_2(K - F(t)), F(0) = 0, \tag{1.22}$$

其中 a_1 和 a_2 是大于零的比例系数. 现在考虑 $N_1(t)$为 t 时刻已购买了新产品的革新者人数，并且注意到这部分人只要接收到搜寻型信息后立即购买，那么有 $N_1(t)=F(t)$. 因此由(1.21)和(1.22)得

$$\frac{\mathrm{d}N_1(t)}{\mathrm{d}t} = (K_1 - N_1(t))(\alpha + \beta N_1(t)), N_1(0) = 0. \tag{1.23}$$

(1.23)式即为关于革新者方程.

由假设 5，模仿者一旦得到了产品使用的体验型信息后决定购买. 而体验型信息是模仿者通过与已购买使用了该产品的人们直接交流得到的，那么此类促使模仿者采取购买行动的信息其传播方式应该是纯口传的. 假设人们在购买产品后需要经过时间 T 才能得出对该产品的评价，并且绝大多数人在购买后感到满意，持不满意见的人数可忽略.

设 $N_2(t)$为 t 时刻购买了新产品的模仿者人数，得到如下关于模仿者方程为

$$\frac{\mathrm{d}N_2(t)}{\mathrm{d}t} = (K_2 - N_2(t))(\gamma_1 N_1(t-T) + \gamma_2 N_2(t-T)) \tag{1.24}$$

显然当 $t\leqslant T$ 时，$N_2(t)=0$，即方程中 $t\geqslant T_0$. $\gamma_1,\gamma_2>0$，表示还未购买新产品模仿者分别从已购买了该产品的革新者和模仿者那里得到的产品体验型信息导致购买行为的系数. 方程(1.24)是具有时滞的微分方程，不便于研究. 我们把方程(1.24)简化. 为此假设：

6. 与产品信息传播相比，购买产品后对产品评价所需时间忽略不计，即 $T=0$.

7. 在体验型信息传播方面，革新者与模仿者的作用相同，即 $\gamma_1=\gamma_2$.

这时微分方程(1.24)化为

$$\frac{\mathrm{d}N_2(t)}{\mathrm{d}t} = \gamma(K_2 - N_2(t))(N_1(t) + N_2(t)), N_2(0) = 0, \tag{1.25}$$

方程(1.23)和(1.25)一起，描述了新产品扩散的过程. 称为 Steffens-Murthy 模型，简记为 S-M 模型. 由 S-M 模型解出 $N_1(t)$和 $N_2(t)$，则 t 时刻新产品购买者总数为

$$N(t) = N_1(t) + N_2(t).$$

由于耐用消费品短期内不会重复购买，故 $N(t)$可以看作是随时间 t 变化的产品累积销售量函数. 把 $N_1(t)$和 $N_2(t)$分别按各时期列出销售量表并绘图，可以看到它们各自呈

单峰结构. 但这两个单峰叠加后则可能产生双峰结构,也可能仍为单峰,这是 S-M 模型的灵活之处.

第五节 作战模型

第一次世界大战期间,F. W. Lanchester 提出了几个尚不成熟的作战模型. 从那以后,人们不断地推广这些模型,并用 Lanchester 模型分析一些战争实例. 本节介绍了三个 Lanchester 模型. 并且介绍了一个二次大战中的著名战役,即美日的硫黄岛之战.

一、三个 Lanchester 战斗模型

一支 x 部队和一支 y 部队互相交战. 设 $x(t)$和 $y(t)$分别代表两个部队在 t 时刻的力量,其中 t 从战斗开始时以天计算. 视力量 $x(t)$和 $y(t)$为士兵的数量,并且假定 $x(t)$和 $y(t)$是连续地变化,并且为时间的可导函数.

对于 x 部队,由于各种不可避免的疾病、开小差以及其它非作战事故所引起的损失率,称为自然损失率(OLR_x);由于与 y 部队遭遇而产生的战斗损失率(CLR_x),以及补充率(RR)等信息. 假定 $x(t)$的净变化率由下式给出

$$\frac{dx(t)}{dt} = -(OLR_x + CLR_x) + RR \tag{1.26}$$

y 部队也有一个类似的方程. 问题是要求出关于每个部队的这些变化率的适当公式,然后分析相应微分方程的解 $x(t)$和 $y(t)$,从而确定谁将"赢得"战斗.

在方程(1.26)中,一般可假设自然损失率 OLR_x 与士兵的人数 x 成正比,即 $OLR_x = ax$,$OLR_y = by$. 而战斗损失率与部队的作战形式有关.

如果 x 部队是正规部队,打的是常规战,x 部队的士兵公开地活动,处于 y 部队每一个士兵的监视和杀伤范围之内,一旦 x 部队某个士兵被杀伤,y 部队的火力立即集中在其余士兵身上,那么 x 部队的战斗损失率 CLR_x 与 y 部队士兵的人数 y 成正比,即 $CLR_x = by$,b 表示 y 部队的战斗效果系数,它可以表示为 $b = r_y p_y$,其中 r_y 是 y 部队的射速(每天每个士兵的射击次数),p_y 表示一次射击杀死一个敌人的可能性.

如果 x 部队打游击战,它的士兵在 y 部队看不到的某个面积为 A_x 的隐蔽区域内活动,y 部队士兵不是向 x 部队士兵开火,而是向面积为 A_x 的区域内射击,并且不知道杀伤情况. 此时 x 部队的战斗损失率不仅与 y 成正比,而且与 x 也成正比. 即 $OLR_x = gxy$. 战斗效果系数 $g = r_y p_y = r_y \dfrac{A_{ry}}{A_x}$,其中 r_y 仍是射击率,命中率 p_y 是 y 部队一次射的有效区域面积 A_{ry}与 x 部队整个活动区域面积 A_x 之比.

在以上假设下,Lanchester 根据战争的不同特性,给出了三个不同的模型:

常规战争

$$\begin{cases} \dfrac{dx(t)}{dt} = -ax(t) - by(t) + F(t) \\ \dfrac{dy(t)}{dt} = -cx(t) - by(t) + G(t) \end{cases}$$

游击战争

$$\begin{cases}\dfrac{dx(t)}{dt} = -ax(t) - gx(t)y(t) + F(t)\\ \dfrac{dy(t)}{dt} = -dy(t) - hx(t)y(t) + G(t)\end{cases}$$

常规-游击战混合型

$$\begin{cases}\dfrac{dx(t)}{dt} = -ax(t) - gx(t)y(t) + F(t)\\ \dfrac{dy(t)}{dt} = -cx(t) - dy(t) + G(t)\end{cases}$$

在这三个模型中，a、b、c、d、h、g 为非负损失率常数，$F(t)$和 $G(t)$为以兵员数量计算的每天补充率.

二、常规战斗模型

在对战争的演变和预测结局作初步分析时，我们考虑双方都没有增援的情况，即 $F(t)=G(t)=0$，并且假设自然损失率为零. 则常规战争模型简化为

$$\begin{cases}\dfrac{dx}{dt} = -by & (1.27)\\ \dfrac{dy}{dt} = -cx & (1.28)\end{cases}$$

用(1.28)除以(1.27)式，可以得到

$$\frac{dy}{dx} = \frac{bx}{ay} \tag{1.29}$$

把(1.29)式分离变量并积分得

$$\begin{gathered} b\int_{y_0}^{y(t)} y\mathrm{d}y = c\int_{x_0}^{x(t)} x\mathrm{d}x \\ b(y^2(t) - y_0^2) = c(x^2(t) - x_0^2) \end{gathered} \tag{1.30}$$

式中 x_0，y_0 为战斗开始时，双方的战斗力，令 $K=ay_0^2-cx_0^2$，则(1.30)式化为

$$by^2 - cx^2 = K. \tag{1.31}$$

(1.31)式确定的图形是双曲线族，图 1-2 绘出了对应不同 K 值的双曲线. 显然，我们只需考察兵力象限($x\geqslant0,y\geqslant0$)中的曲线. 曲线上的箭头表示兵力随时间而变的方向. 因为 $x(t)>0$，$y(t)>0$，就有$\frac{dx}{dt}<0$ 和$\frac{dy}{dt}<0$. 所以箭头的方向如图 1-2 所示.

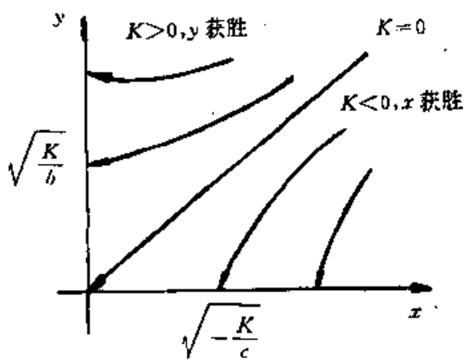


图 1-2

如果 $K>0$，轨线与 y 轴交于$\sqrt{\frac{K}{b}}$，即存在 t_1，当 $x(t_1)=0$ 时，$y(t_1)=\sqrt{\frac{K}{b}}>0$，这表明 y 部队获胜，同理可知，当 $K=0$ 时，双方战平，若 $K<0$，则 x 部队获胜. 如果 y 部

队取胜，就必须形成一个 $K>0$ 的战斗态势，$K>0$ 等价于 $\left(\frac{y_0}{x_0}\right)^2>\frac{b}{a}$. 注意 $b=r_yp_y$，$c=r_xp_x$，y 部队取胜的条件可表示为

$$\left(\frac{y_0}{x_0}\right)^2>\frac{r_xp_x}{r_yp_y} \tag{1.32}$$

这表明初始时刻兵力之比对战争的结局是以平方关系起作用的，即如果 y 部队投入的初始兵力从 $\frac{y_0}{x_0}=1$ 增加到 $\frac{y_0}{x_0}=2$，那么 y 部队影响战争的结局的能力从 1 增加到 4. 正是由于这个缘故，常规战争的这种规律称为平方律.

方程(1.31)只与两支部队各自的兵力有关，而与时间的推移无关. 通过下述方法，可得到关于兵力的瞬时变化公式，对(1.27)式微分，并利用(1.28)式可得

$$\frac{d^2x}{dt^2}=-b\frac{dy}{dt}=-bcx$$

则 $x(t)$ 满足初值问题

$$\begin{cases}\frac{d^2x}{dt^2}-bcx=0,\\ x(0)=x_0,x'(0)=-by_0\end{cases}$$

的解为

$$x(t)=x_0\cos\beta t-\gamma y_0\sin\beta t \tag{1.33}$$

其中 $\beta=\sqrt{bc}$，$\gamma=\sqrt{\frac{b}{c}}$，类似地，

$$y(t)=y_0\cos\beta t-\frac{x_0}{\gamma}\sin\beta t \tag{1.34}$$

图 1-3，给 $K>0$ 的特殊情况下，(1.33)式和(1.34)式的图形. 可以看出，y 部队要取胜并不一定要求 y_0 超过 x_0，但必须使 $by_0^2>cx_0^2$ 或等价地 $\gamma y_0>x_0$.

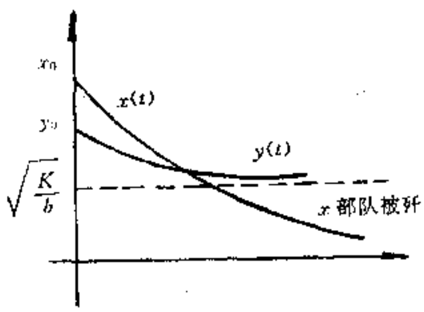


图 1-3

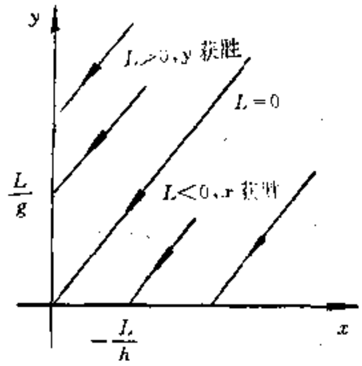


图 1-4

三、游击战模型

假设两支游击部队在作战中均无自然损失和增援，在这些严格条件下，游击战模型将化为

$$\begin{cases}\dfrac{\mathrm{d}x}{\mathrm{d}t}=-gxy\\[2mm]\dfrac{\mathrm{d}y}{\mathrm{d}t}=-hxy\end{cases}$$

消去 $\mathrm{d}t$ 得

$$\frac{\mathrm{d}y}{\mathrm{d}x}=\frac{h}{g} \tag{1.35}$$

解为

$$g(y(t)-y_0)=h(x(t)-x_0) \tag{1.36}$$

设 $L=gy_0-hx_0$，则由(1.35)得到线性律

$$gy-hx=L. \tag{1.37}$$

如果 $L>0$，即 $\frac{y_0}{x_0}>\frac{h}{g}$，则 y 部队将获胜．同样 $L<0$ 时，x 部队获胜．(1.37)式的图形如图1-4，由于(1.37)式的线性形式，即游击战中 $\frac{y_0}{x_0}$ 只起一次方作用，这种规律称为线性律．

四、常规-游击战混合模型

此时，一支游击部队与一支常规部队交战．假设无增援也没有自然损失．在这种情况下，模型为

$$\begin{cases}\dfrac{\mathrm{d}x}{\mathrm{d}t}=-gxy\\[2mm]\dfrac{\mathrm{d}y}{\mathrm{d}t}=-cx\end{cases}$$

其中 x 表示游击部队，y 表示常规部队，消去 $\mathrm{d}t$ 得

$$\frac{\mathrm{d}y}{\mathrm{d}x}=\frac{c}{gy}.$$

对其积分，得抛物律

$$gy^2(t)=2cx(t)+M \tag{1.38}$$

其中 $M=gy_0^2-2cx_0$，如果 $M<0$，那么游击队获胜，而 $M>0$ 时，常规部队获胜．(1.38)确定的图形是抛物线族，如图1-5．若正规 y 部队获胜，则 $M>0$，或 $\left(\frac{y_0}{x_0}\right)^2>\frac{2c}{gx_0}$，以 $b=r_xp_x$，$g=r_y\frac{A_{ry}}{A_x}$ 代入，得

$$\left(\frac{y_0}{x_0}\right)^2>2\frac{r_x}{r_y}\frac{A_xp_x}{A_{ry}}\cdot\frac{1}{x_0} \tag{1.39}$$

下面利用(1.39)式分析一下以正规部队作战的 y 为了战胜游击部队 x 需投入多大的兵力 y_0．设甲方兵力 $x_0=100$，命中率 $p_x=0.1$，射击率 r_x 是 y 部队 r_y 的一半，整个活动区域的面积是0.1平方千米，y 部队每次射击的有效面积是1平方米，那么 y 部队取得胜利必须使

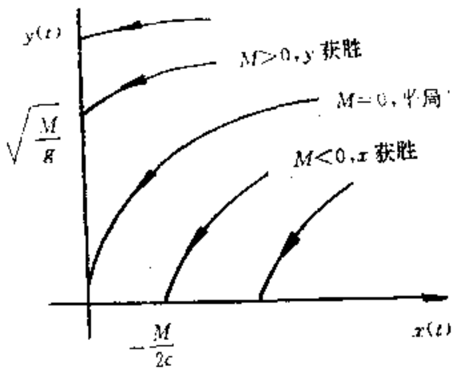


图 1-5

$$\left(\frac{y_0}{x_0}\right)^2 > 2 \cdot \frac{1}{2} \frac{0.1 \times 10^6 \times 0.1}{1 \times 100} = 100,$$

即$\frac{y_0}{x_0}>10$，y部队必须10倍于游击部队x的兵力.

在越南战争中，到1968年春，美国在越南战场投入的兵力约为168万，越南北越和越共的兵力约为28万，兵力比为

$$\frac{1\ 680\ 000}{280\ 000} = 6$$

在1968年2月间的新年攻势刚过，Westmoreland将军就向美国总统约翰逊要求增派一支20.6万人的部队.如果约翰逊总统根据要求派出20.6万人的部队，那么，双方力量对比将增加到

$$\frac{1\ 886\ 000}{280\ 000} = 6.7$$

这仍不足以使常规部队的状况有较大的改变.而越南部队也将会增加到31.4万人，比例仍保持6∶1.正是基于这样的分析，以及美国人民对整个事态的焦虑，约翰逊总统才不得不从政治上寻求一种解决越南问题的办法.他拒绝了Westmoreland的要求，发起了巴黎和平会谈，最终美国于1973年撤离战斗，越南取得了最后的胜利.

五、硫黄岛战斗模型

硫黄岛位于东京以南660英里，是一座火山岛，面积仅8平方英里.二次世界大战中，最残酷的一场战斗就是在这个岛上进行的，交战国是日本和美国.硫黄岛作为日军的重要基地，可以组织攻击前去轰炸东京和日本其它城市的美国飞机，相反美国要占领它作为自己紧靠日本的轰炸机基地.美军于1945年2月19日开始袭击，日军要求不惜一切代价守住这个岛，只要还有一个人，就要进行下去.美军志在必得.激烈的战斗持续了一个月，双方伤亡惨重，日方守军21 500人，其中20 000多人阵亡，其余被俘，美军共投入73 000人，伤亡20 265人.战争进行到第28天时美军宣布占领了这个岛，实际战斗到36天才结束.

J. H. Engel利用美军战地记录构造了硫黄岛战斗的Lanchester模型.分别用$A(t)$和$J(t)$表示t天的美军和日军士兵人数.在战争期间，美军得到了支援，而日军没有.设美军的补充率为$F(t)$.忽略非战斗减员，即认为自然损失率为零，则硫黄岛战斗模型为

$$\frac{\mathrm{d}A}{\mathrm{d}t} = -bJ(t) + F(t) \tag{1.40}$$

$$\frac{\mathrm{d}J}{\mathrm{d}t} = -cA(t) \tag{1.41}$$

对(1.41)微分，然后利用(1.40)得

$$\frac{\mathrm{d}^2 J}{\mathrm{d}t^2} - bcJ = -cF(t) \tag{1.42}$$

方程(1.42)是一个带有“驱动力”$-cF(t)$的二阶常系数非齐次线性微分方程，可以解得

$$J(t) = J_0\cos(\beta t) - \frac{A_0}{\gamma}\sin(\beta t) - \frac{1}{\gamma}\int_0^t \sin\beta(t-s)F(s)\mathrm{d}s \tag{1.43}$$

其中 J_0 和 A_0 分别为袭击前岛上日军和美军的兵力，$\beta=\sqrt{bc}$，$\gamma=\sqrt{\frac{b}{c}}$. A_0 出现在(1.43)式是因为 $\left.\frac{\mathrm{d}J}{\mathrm{d}t}\right|_{t=0}=-cA_0$.

将(1.43)式对 t 微分，然后利用(1.41)就可以确定 $A(t)$，由对积分式求微分的莱布尼兹法则

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_0^t h(t,s)\mathrm{d}s = h(t,t) + \int_0^t \frac{\partial}{\partial t}h(t,s)\mathrm{d}s \tag{1.44}$$

对(1.43)微分，注意 $\sin 0=0$，$\frac{\beta}{r}=c$，得

$$\frac{\mathrm{d}J}{\mathrm{d}t} = \beta J_0\sin(\beta t) - cA_0\cos(\beta t) - c\int_0^t \cos\beta(t-s)F(s)\mathrm{d}s$$

因为 $A(0)=A_0=0$，并利用(1.41)中 $A(t)=-\frac{1}{c}\frac{\mathrm{d}J}{\mathrm{d}t}$，我们可以得到两军兵力随时间变化的规律

$$A(t) = -\gamma J_0\sin(\beta t) + \int_0^t \cos\beta(t-s)F(s)\mathrm{d}s \tag{1.45}$$

$$J(t) = J_0\cos(\beta t) - \frac{1}{\gamma}\int_0^t \sin\beta(t-s)F(s)\mathrm{d}s \tag{1.46}$$

初始条件为

$$A(0) = 0, J(0) = J_0 = 21\ 500,$$

美国的增援率 $F(t)$ 为

$$F(t) = \begin{cases} 54\ 000 & 0 \leqslant t < 1 \\ 0 & 1 \leqslant t < 2 \\ 6\ 000 & 2 \leqslant t < 3 \\ 0 & 3 \leqslant t < 5 \\ 13\ 000 & 5 \leqslant t < 6 \\ 0 & 6 \leqslant t \leqslant 36 \end{cases}$$

现在就剩下 β 和 γ 是待定的了. 为此得确定 b 和 c. b 和 c 不能仅仅是现役战斗员而应是整个部队的平均战斗效果系数. 设美军在 $t=1,2,\cdots,36$ 时的实效人数为 $A_{act}(t)$. 为了估计系数 c，对(1.41)式作积分，并用求和代替，得

$$J(36) - J(0) = -c\int_0^{36} A_{act}(t)\mathrm{d}t = -c\sum_{t=1}^{36} A_{act}(t)$$

由战地记录得 $\sum_{t=1}^{36} A_{act}(t)=2\ 037\ 000$，于是

$$c = \frac{J(0) - J(36)}{\sum_{t=1}^{36} A_{act}(t)} = \frac{21\ 500 - 0}{203\ 700} \approx 0.010\ 6 \tag{1.47}$$

利用(1.47)给出的 c 值，我们可以求出日军在战时每天的实际人数 $J_{act}(t)$ 的近似值.

$$J_{act}(t) \approx J(0) - c\sum_{k=1}^{t} A_{act}(k),$$

$$= 21\,500 - 0.010\,6\sum_{k=1}^{t} A_{act}(k), t = 0,1,\cdots,36. \quad (1.48)$$

这里，我们对(1.41)式求从 0 到 t 的积分，然后用(1.48)中的和代换了 $\int_0^t A_{act}(s)ds$. 由于第 28 天到第 36 天交战，只是零星地进行，为求出 b，我们令 $t=28$.（在求 c 时，选择 $t=36$，是因为只有当整个战斗结束时，我们才能获得准确的日军军队损失数）对(1.40)积分

$$A(t) = A_0 - b\int_0^t J(s)ds + \int_0^t F(s)ds$$
$$\approx - b\sum_{k=1}^{t} J_{act}(k) + \sum_{k=1}^{t} F(k)$$

因为 $\sum_{k=1}^{28} F(k)=73\,000$，由(1.48)式可算出 $\sum_{k=1}^{28} J_{act}(t)=372\,500$，根据美军伤亡数可得 $A(28)=73\,000-20\,165=52\,735$，所以

$$b \approx \frac{\sum_{k=1}^{28} F(k) - A(28)}{\sum_{k=1}^{28} J_{act}(k)} = 0.054\,4$$

从而得到 $\beta=0.024\,013$、$\gamma=2.265\,409$，由(1.45)和(1.46)式可以得到每天美军和日军(理论的)士兵人数

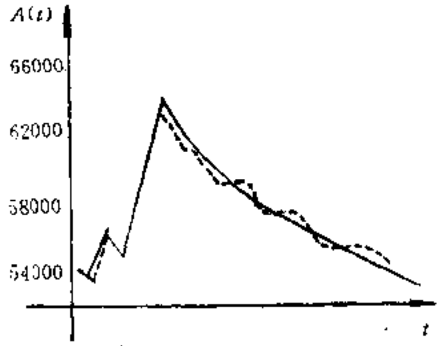


图 1-6

$$A(t) = -48\,706.293\,5\sin 0.024\,013t + \int_0^t \cos 0.024\,013(t-s)F(s)ds,$$

$$J(t) = 21\,500\cos 0.024\,013t - 0.441\,422\int_0^t \sin 0.024\,013(t-s)F(s)ds.$$

在图 1-6 中，给出了美军战地记录的实际数据与模型的理论结果进行了比较，可以看出二者吻合得相当好.

第六节　种群生态学模型

生态学是生物学中研究生物的宏观现象的科学，研究生物体与环境之间的关系. 研究生物群体之间的关系. 研究过程中，这些关系往往可以通过一个数学模型来描述，这样的数学模型就称为种群生态学模型. 在种群生态学中，目的是要想办法去建立能够比较准确地描述这个系统的数学模型，然后通过数学的计算和数学理论分析去解释，去了解一些生态现象，以达到人类对某些生态现象的控制.

一、单种群模型

我们假设所研究的种群密度分布是均匀的，因此种群的密度可以表示成时间 t 的函数 $N(t)$，假定种群密度随时间 t 的变化率与当时种群密度成正比，即模型为

$$\frac{dN}{dt} = \gamma N \tag{1.49}$$

这里 r 称为种群的内禀自然增长率(出生率减去死亡率). $N(t)$满足初值 $N(0)=N_0$ 的解为

$$N(t) = N_0 e^{rt} \tag{1.50}$$

容易看出,当 $t\to\infty$时,有 $N(t)\to\infty$. 这也就说明模型(1.49)只能对较短时间的情况实用,因为在自然界中,种群的密度不可能增长到无限,而是有一个最大的限度 K,1938 年 Verhulst 提出把模型(1.49)修改为

$$\frac{dN}{dt} = rN(1-\frac{N}{K}) \tag{1.51}$$

其中 K 为环境容纳量,(1.51)式满足初值条件 $N(0)=N_0$ 的解为

$$N(t) = K[1+(\frac{K}{N_0}-1)\exp(-rt)]^{-1} \tag{1.52}$$

我们可以看出(1.51)式有两个平衡点,其一是 $N=0$,另一个是 $N=K$(当初值取到 $N_0=0$ 或 $N_0=K$ 时,有$\frac{dN}{dt}=0$). 平衡点 $N=0$ 是不稳定的. 这是因为任何接近于 $N=0$ 的初值出发的(1.51)的解当 t 增加时都远离 $N=0$. 而平衡点 $N=K$ 是渐近稳定的. 渐近稳定是指对于任意给定的 $\varepsilon>0$,必存在 $\delta>0$,使当初值在$(K-\delta, K+\delta)$内的任意解 $N(t)$均有 $|K-N(t)|<\varepsilon$,且当 $t\to\infty$时,$N(t)\to K$,见图 1-7. 这就说明由模型(1.51)所描述的生态现象必将达到生态平衡. 种群不会绝灭,也不会无限繁殖.

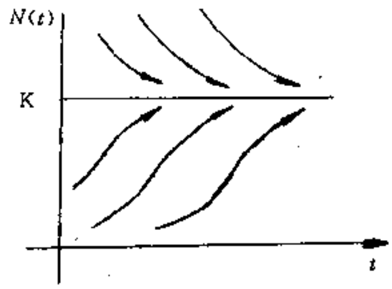


图 1-7

方程(1.51)被称为 Logistic 模型,它比指数增长模型(1.49)更接近于实际,但它也有缺点. 有时会产生较大偏差. 人们认为方程产生偏差的原因是对密度制约效应的线性化,我们把方程(1.51)写为

$$\frac{1}{N(t)}\frac{dN(t)}{dt} = r - \frac{r}{K}N(t)$$

上式右端是 $N(t)$的线性函数,若更为接近实际情况,则应该用非线性密度制约函数,即方程改写为

$$\frac{dN}{dt} = NF(N) \tag{1.53}$$

这就是单种群的一般模型. 关于模型(1.53),我们希望了解是否和(1.51)一样,也存在一个稳定的平衡态,而且是全局渐近稳定的,也就是说,不管种群密度的初始数量如何,经过适当长的时间后,种群密度将保持在平衡态的数量,为此我们给出如下的定理.

定理 1　如果函数 $F(N)$满足下列条件,则模型(1.53)的正平衡点是全局渐近稳定的.

(1)有一个正的平衡位置 N^*,即存在 $N^*>0$,使 $F(N^*)=0$;

(2)对于所有的 N,若 $N^*>N>0$,有 $F(N)>0$,且

(3)对于所有的 N,若 $N>N^*$,有 $F(N)<0$.

例 3 Gilpin 和 Ayala 考虑如下种群生态学模型

$$\frac{dN}{dt}=rN\left[1-\left(\frac{N}{K}\right)^\theta\right],$$

其中 r、K 和 θ 是正常数,有平衡点 $N=K$,这里 $F(N)=r\left[1-\left(\frac{N}{K}\right)^\theta\right]$,当 $0<N<K$ 时,$F(N)>0$;当 $N>K$ 时,$F(N)<0$. 由定理 1,平衡位置 $N=K$ 是全局稳定的.

模型(1.51),(1.53)是描述种群在自然环境下增长规律的模型,这些种群没有受到人类的开发. 例如在渔业中,方程(1.53)只描述鱼在自然环境下生长的情况,没有考虑到人类的捕捞,如果把人类的捕捞因素考虑进去,则模型就要做相应的修改. 这时具有收获率的单种群模型为

$$\frac{dN}{dt}=NF(N)-u(t), \tag{1.54}$$

$$\frac{dN}{dt}=NF(N)-u(t)N, \tag{1.55}$$

在渔业生产中,我们知道,并不是一年中把鱼都捞干净,鱼的产量最高,而是要考虑怎样控制每年的鱼产量,才能有利于鱼的繁殖,使得在一定的时间内,例如 10 年、20 年,鱼的产量最高. 我们用模型(1.54)来研究,则渔业生产中的这个问题的数学描述为

状态方程:$\frac{dN}{dt}=NF(N)-u(t)$,

初值:$N(0)=N_0$,

终值:$T=\text{const}$,$N(T)\geqslant a$,

条件:$0\leqslant u(t)\leqslant u_{max}$,

目标:$\max\int_0^T u(t)dt$.

这里 T 就是上面所说的 10 年,20 年. $N(T)\geqslant a$,即表示要求鱼的数量仍保持一定. 这里我们是以总捕获量最大为目标的. 我们可以利用最优化方法中极大值原理解决这个问题.

二、简单的两种群相互作用的模型

本世纪 20 年代中期,意大利生物学家 D'Ancona 研究了相互制约的各种鱼类群体的变化情况. 在研究过程中,他偶然注意到第一次世界大战期间,在原南斯拉夫的里耶卡港,人们捕获的鱼类中,鲨鱼、鳐鱼等软骨鱼的百分比有大幅度增加见表 1-2,而这类鱼并不象食用鱼那么合乎人们的需要. 软骨鱼和食用鱼的区别在于,软骨鱼是捕食者,而食用鱼是软骨鱼的捕获对象,即食饵. 软骨鱼依靠食用鱼而生存. D'Ancona 以为由于在战争期间捕鱼的强度大大地降低了,软骨鱼就可以得到更多的捕获物了,因此它也就快速地增殖和茁壮地成长起来. 但他无法解释在这期间食用鱼也更丰富了. D'Ancona 的分析仅仅表明,当捕鱼水平降低时,可以得到更多的软骨鱼;但不能解释为什么降低捕鱼水平时,对软骨鱼与食用鱼比较而言,更有利于软骨鱼. 百思不解的 D'Ancona 求助于他的同事著名的意大利数学家 Volterra,希望他能对软骨鱼及食用鱼的增长情况建立一个数学模型. Volterra 成功地利用微分方程模型回答了这个问题.

表 1-2

年 代	1914	1915	1916	1917	1918	1919	1929	1921	1922	1923
百分比	11.9	21.4	22.1	21.2	36.4	27.3	16.0	15.9	14.8	10.7

Volterra 把所有的鱼分成两类，即食饵(食用鱼)和捕食者(软骨鱼). t 时刻食饵的数量为 $x(t)$，捕食者的数量为 $y(t)$.

当只考虑食饵，不考虑捕食者时，假定食饵的增长遵循如下规律.

$$\frac{\mathrm{d}x}{\mathrm{d}t} = Ax$$

考虑捕食者的存在，食饵的生长将减缓其速度，此时食用鱼的增长满足方程

$$\frac{\mathrm{d}x}{\mathrm{d}t} = Ax - Bxy$$

捕食者的增长可作类似的讨论，因此得 Volterra 方程

$$\begin{cases} \dfrac{\mathrm{d}x}{\mathrm{d}t} = Ax - Bxy & (1.56) \\ \dfrac{\mathrm{d}y}{\mathrm{d}t} = cxy - Dy & (1.57) \end{cases}$$

其中 A、B、C、D 均为大于零的常数. 我们在第一象限内进行讨论.

首先令 $\frac{\mathrm{d}x}{\mathrm{d}t}=0, \frac{\mathrm{d}y}{\mathrm{d}t}=0$ 求出平衡点，因为

$$\begin{cases} Ax - Bxy = 0 \\ Cxy - Dy = 0 \end{cases}$$

的解为 $x_0=0, y_0=0; x^*=\frac{D}{C}, y^*=\frac{A}{B}$. 所以 Volterra 方程有两个平衡点 $O(0,0)$，$P(x^*, y^*)=(\frac{D}{C}, \frac{A}{B})$. 我们称

$$x(t) = \frac{D}{C}, y(t) = \frac{A}{B}$$

为 Volterra 方程的平衡解. 显然 x 轴与 y 轴都是它的解曲线. 由解的唯一性可知，对 $t=t_0$ 由第一象限 $x>0, y>0$ 出发的每一个解 $x(t), y(t)$；当 $t \geqslant t_0$ 时都保持在第一象限内.

由 $\frac{\mathrm{d}x}{\mathrm{d}t}=0$，得到的 $y=\frac{A}{B}$ 和 $\frac{\mathrm{d}y}{\mathrm{d}t}=0$ 得到的 $x=\frac{D}{C}$ 把第一象限分成四块，每一块中 $\frac{\mathrm{d}x}{\mathrm{d}t}$ 与 $\frac{\mathrm{d}y}{\mathrm{d}t}$ 的符号不变. 方向大致如图 1-8 所示. 将(1.56)与(1.57)两式相除，得方程

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{(Cx - D)y}{(A - By)x}.$$

分离变量，求出解

$$Cx + By - D\ln x - A\ln y = k$$

也就是说，Volterra 方程的轨线是沿着函数

$$H(x,y) = Cx + By - D\ln x - A\ln y$$

的等位线. 显然，函数 $H(x,y)$ 在 $x \geqslant 0, y \geqslant 0$ 内的点 $P(x^*, y^*)=\left(\frac{D}{C}, \frac{A}{B}\right)$ 为唯一的极值

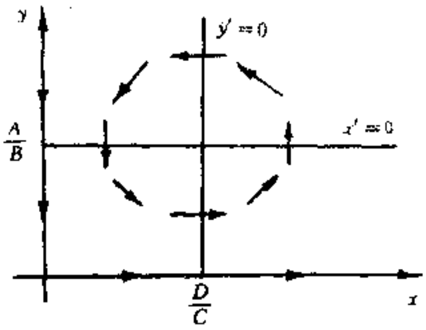


图 1-8

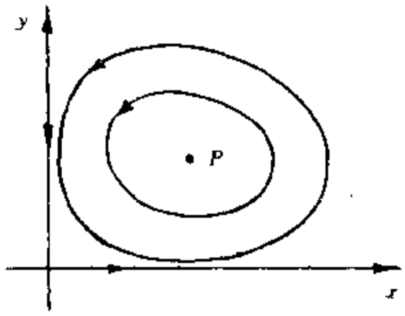


图 1-9

点(极小).所以方程的每一条轨线(除去平衡点与坐标轴之外)都是闭轨. P 是中心,如图 1-9.

当 $x(0)$和 $y(0)$是正的时,Volterra 方程的一切解 $x(t)$,$y(t)$都是时间 t 的周期函数.即 $x(t+T)=x(t)$,$y(t+T)=y(t)$,其中 T 为某个正数,D'Ancona 数据是捕食者每年的年平均数.为了比较,我们必须算出 Volterra 方程的解 $x(t)$,$y(t)$的平均值.我们易算出

$$\bar{x}=\frac{1}{T}\int_0^T x(t)\mathrm{d}t=\frac{1}{T}\int_0^T\left(\frac{D+y'/y}{C}\right)\mathrm{d}t$$

$$=\frac{D}{C}+\frac{1}{CT}[\ln y(T)-\ln y(0)]$$

由于 $y(t)$的周期性,$y(0)=y(T)$,故上式得 $\bar{x}=\frac{D}{C}$,同理可得 $\bar{y}=\frac{A}{B}$,由此可见,$x(t)$,$y(t)$的平均值就是平衡值.

从图 1-9 可见,如果开始时只有捕食者而没有食饵,结果是捕食者死尽;如果从一开始就没有捕食者,那么食饵会无限增长;如果开始时捕食者数为$\frac{A}{B}$,食饵数为$\frac{A}{D}$,则将永远维持这个平衡状态;如果开始时两种群初值 $x(0)>0$,$y(0)>0$,但不是$\left(\frac{D}{C},\frac{A}{B}\right)$,则捕食者与食饵的数将循环振荡,两种群都不会绝灭,也不会无限增长.

现在,我们考虑捕鱼的影响.捕鱼使食用鱼的总数按 $\varepsilon x(t)$的速度减少,软骨鱼总数按 $\varepsilon y(t)$的速度减少.常数 ε 反映着捕鱼的强度;即与海上的船只数目和水中网的数目有关.这时 Volterra 方程改变为

$$\frac{\mathrm{d}x}{\mathrm{d}t}=Ax-Bxy-\varepsilon x=(A-\varepsilon)x-Bxy \tag{1.58}$$

$$\frac{\mathrm{d}y}{\mathrm{d}t}=Cxy-Dy-\varepsilon y=Cxy-(D+\varepsilon)y \tag{1.59}$$

这里 $A-\varepsilon>0$.类似上面的推导,我们易得方程(1.58)和(1.59)的平均值为

$$\bar{x}=\frac{D+\varepsilon}{C},\bar{y}=\frac{A-\varepsilon}{B}$$

因此,适当地增加捕鱼量($\varepsilon<A$),将使食用鱼的数量增加,而使鲨鱼等软骨鱼的数量减少.相反,降低了捕鱼水平,那么,按平均数来计算,软骨鱼的数量在增加,而食用鱼的数量在减少.这就是 Volterra 原理.

三、数学实验

Pearl 与 Read 曾研究了美国人口的增长问题.他们给出了如下预测方程

$$N(t)=\frac{197\ 268\ 200}{1+e^{-0.031\ 34(t-1\ 914.3)}}$$

我们把模型预报值与实际数据列在下面：

表 1-3

年	实 际	预 报	误 差	百分比
1790	3 929 000	3 929 000	0	0.0
1800	5 308 000	5 336 000	28 000	0.5
1810	7 240 000	7 228 000	−12 00	−0.2
1820	9 638 000	9 757 000	119 000	1.2
1830	12 866 000	13 109 000	243 000	1.9
1840	17 069 000	17 506 000	437 000	2.6
1850	23 192 000	23 192 000	0	0.0
1860	31 443 000	30 412 000	−1 031 000	−3.3
1870	38 558 000	39 372 000	814 000	2.1
1880	50 156 000	50 177 000	21 000	0.0
1890	62 948 000	62 769 000	−179 000	−0.3
1900	75 995 000	76 870 000	875 000	1.2
1910	91 972 000	91 972 000	0	0.0
1920	105 711 000	107 559 000	1 848 000	1.7
1930	122 775 000	123 124 000	349 000	0.3
1940	131 669 000	136 653 000	4 984 000	3.8
1950	150 697 000	149 053 000	−1 644 000	−1.1
1960	173 300 000	158 800 000	−1 450 000	−0.9
1970	204 000 000	168 600 000	−35 400 000	−17.4

33

问题 1　对表 1-3 中的数据，用 Mathemetica 软件，给出 Logistic 微分方程的模型

$$\frac{dN}{dt}=rN(1-\frac{N}{K})$$

的参数 r 与 N.并与表 1-3 中的预报值比较，您的预报值是否更好？

问题 2　对表 1-3 中的数据，用 Mathemtica 软件，给出方程

$$\frac{dN}{dt}=-\alpha N\ln\frac{N}{K}$$

的参数 α 与 K，并比较此模型的预报值与其它模型的预报值哪一个更好？分析原因.

第七节 交通流模型

60年代以来，由于汽车社会拥有量，越来越多．伴随而来的交通问题也越来越多．因此交通问题引起了人们的广泛关注．人们试图分析和预测公路上的交通流量，从而形成了交通流理论．通常交通流模型可以分为两类．第一类模型是宏观模型．它实际上是把交通流量当作液体的流量来处理，它涉及到整个交通线总的或平均的变化，另一类是微观模型，在微观模型中，研究的问题是：在一条交通线中，单个汽车的相互作用，具体地说，可以研究跟随汽车的加速度，此加速度是领头者和跟随者之间的距离、两汽车的相对速度以及跟随汽车驾驶员的反应时间的函数．这些模型称为汽车的跟随模型．本节仅对宏观交通流模型作简单介绍．

一、交通流的连续性假设

我们的研究对象是宏观的公路上车辆行驶的状况，考虑一般公路行驶的若干个汽车的流动问题．我们利用类似于在流体力学中分析一个管道中液体的流动的理论来分析这个问题．我们先给出连续性假设．我们把公路上行驶的一辆接一辆的汽车看作是如同水流样的连续流．之所以这样假设，是因为公路的长度远远大于汽车之间的距离．

选公路为 x 轴，轴的正方向指向车流的运动方向．令 $n(x,x+\Delta x,t)$ 表示 t 时刻点 x 与点 $x+\Delta x$ 之间的汽车数量．我们假设存在 $\rho(x,t)$，使得对任意 $x,\Delta x$ 和 t，有

$$n(x,x+\Delta x,t)=\int_x^{x+\Delta x}\rho(y,t)\mathrm{d}y$$

由微积分基本定理，如果 ρ 是连续的，则

$$\rho(x,t)=\lim_{\Delta x\to 0}\frac{n(x,x+\Delta x,t)}{\Delta x}$$

$\rho(x,t)$ 表示在 t 时刻地点 x 处汽车密度．

流量 $q(x,t)$ 就是单位时间内通过 x 点处的汽车数量．因此，在 t 到 $t+\Delta t$ 时间经过点 x 处的总数量为

$$Q(x,t,t+\Delta t)=\int_t^{t+\Delta t}q(x,\tau)\mathrm{d}\tau$$

根据微积分基本定理，有

$$q(x,t)=\lim_{\Delta x\to 0}\frac{Q(x,t,t+\Delta t)}{\Delta t}$$

二、交通流的守恒方程

现在，我们考虑这段路上车辆的平衡或守恒问题，让我们仅考察点 x 到 $x+\Delta x$ 这一段路上的汽车数量的改变率，这里假定汽车行驶时不会出现赶上或超车情况，这段公路上也没有汽车从其它岔路进入或驶出公路．这样车辆守恒就要求在 x 到 $x+\Delta x$ 之间汽车数量的增加率等于这段公路上汽车的流入率减去流出率，设 t 时刻在给定路段 x 到 $x+\Delta x$

之间的车数为 $N(t)$，则从守恒的角度出发我们有

$$N(t+\Delta t) - N(t) = q(x,t)\Delta t - q(x+\Delta x,t)\Delta t$$

上式中令 $\Delta t \to 0$ 得

$$\frac{\mathrm{d}N}{\mathrm{d}t} = q(x,t) - q(x+\Delta x,t) \tag{1.60}$$

另一方面，我们知道，在该区间的汽车总数是整个区间长度上密度 $\rho(x,t)$ 的积分

$$N(t) = \int_x^{x+\Delta x} \rho(y,t)\mathrm{d}y \tag{1.61}$$

于是

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_x^{x+\Delta x} \rho(y,t)\mathrm{d}y = q(x,t) - q(x+\Delta x,t) = -\int_x^{x+\Delta x} \frac{\partial q}{\partial x}\mathrm{d}x \tag{1.62}$$

利用积分中值定值，得到微分形式的守恒律

$$\frac{\partial \rho(x,t)}{\partial t} + \frac{\partial q(x,t)}{\partial x} = 0 \tag{1.63}$$

在速度场中，t 时刻每一点沿 x 轴被给定一个特定的速度记为 $v(x,t)$. 在一般情况下变量 ρ、v 和 q 都是公路坐标 x 和时间 t 的函数，它们三者的关系是

$$q(x,t) = \rho(x,t)v(x,t) \tag{1.64}$$

将(1.64)代入(1.63)可以把问题化为只有两个未知数的一个方程

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x}(\rho v) = 0 \tag{1.65}$$

现在，我们必须设法使路上的汽车速度与车辆密度联系起来，以便求解方程(1.65). 我们假定：一个车辆的速度只取决于交流密度，即

$$v = v(\rho) \tag{1.66}$$

如果在路上没有或有很少的其它车辆，我们希望驾驶员保持最快的速度 v_{max}；然而，当密度增加时，就期望速度减小. 最后，达到某一最大密度时交通就漫漫停滞下来. 因此当 $\rho=\rho_{max}$ 时，$v=0$. 为了便于理解，我们假定一个线性的速度——密度关系，即

$$v = v_{max}\left(1 - \frac{\rho}{\rho_{max}}\right) \tag{1.67}$$

由此可得 Greenshields 模型.

$$q = v_{max}\left(\rho - \frac{\rho^2}{\rho_{max}}\right) \tag{1.68}$$

由此可见，最大流量就是曲线(1.68)上斜率 $\frac{\mathrm{d}q}{\mathrm{d}\rho}$ 为零的那点，由于

$$\frac{\mathrm{d}q}{\mathrm{d}\rho} = v_{max}\left(1 - \frac{2\rho}{\rho_{max}}\right) \tag{1.69}$$

所以，$\rho=\frac{\rho_{max}}{2}$ 时，q 取得最大值，最大值为

$$q_{max} = \frac{1}{4}v_{max}q_{max} \tag{1.70}$$

由方程(1.67)给出的线性关系在纽约城的几条隧道里所进行的研究中已被证明是良好的.

第八节　扩散模型

对于物质的浓度来说，由于物质浓度分布不均匀，物质会由浓度高的地方向浓度低的地方转移，这种现象称为扩散，人们在生活中对扩散现象是比较熟悉的. 如打开一瓶好酒，酒香将会在空气中扩散. 环境保护中的大气污染的扩散. 战争中，炮弹在空中爆炸，烟雾将以爆炸点为中心向外扩散.

在扩散过程中，物质在某点的浓度一般是随着时间 t 和该点的位置 (x,y,z) 的变化而变化，故浓度函数为 $u=u(x,y,z,t)$. 我们以浓度函数作为扩散过程的状态变量，建立数学模型.

我们用 q 表示扩散流强度，它表示单位时间内通过单位横截面积物质的数量. 由实验可知，通过一个曲面的扩散流强度与浓度函数的法向导数成正比. 这个结果称为扩散定理，即

$$q=-k\frac{\partial u}{\partial n} \tag{1.71}$$

其中 k 为扩散系数，负号表示由浓度高向浓度低的地方扩散，下面将以扩散定理为依据建立两个扩散数学模型.

一、一维扩散模型

在制作半导体器件中，某种硅片里需要含某种物质，将这种物质涂在硅片表面放入扩散炉中，这种物质几乎只沿着表面的法线方向扩散到硅片里，所以这是一维扩散问题.

在进行的扩散空间里任何一长方体微小元素. 假定只沿 x 轴方向扩散，由式(1.71)式得

$$q=-\frac{\partial u}{\partial x} \tag{1.72}$$

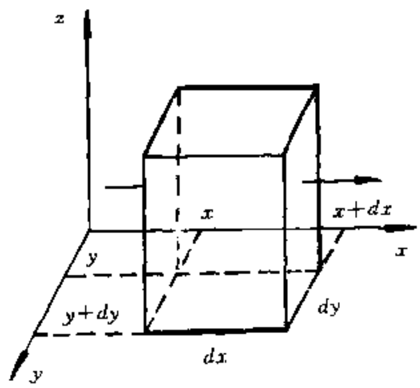


图 1-10

单位时间内由左面流入微小单元体的物质数量为 $q_x\cdot \mathrm{d}y\mathrm{d}z$，从右面流出的量为 $q_{x+\mathrm{d}x}\cdot \mathrm{d}y\mathrm{d}z$. 由于只沿 x 方向扩散，所以前、后、上、下四个面没有物质的扩散，故单位时间内流入微小单元体物质的净数量为

$$\Delta m=(q_x-q_{x+\mathrm{d}x})\cdot \mathrm{d}y\mathrm{d}z=-\frac{\partial q}{\partial x}\mathrm{d}x\mathrm{d}y\mathrm{d}z$$

$$=\frac{\partial}{\partial x}\left(k\frac{\partial u}{\partial x}\right)\mathrm{d}x\mathrm{d}y\mathrm{d}z$$

于是浓度对时间 t 的变化率为

$$\frac{\partial u}{\partial t}=\frac{\Delta m}{\mathrm{d}x\mathrm{d}y\mathrm{d}z}=\frac{\partial}{\partial x}\left(\lambda\frac{\partial u}{\partial x}\right) \tag{1.73}$$

这就是一维扩散方程. 若扩散是均匀的，则 k 等于常数，设 $a^2=k$，则(1.73)式可以写为

$$u_t - a^2 u_{xx} = 0$$

显然扩散方程与热传导方程有相同的形式,因此它的解可以参阅热传方程的解法来得到.

二、大气污染的扩散

为了预防和治理大气污染,需要计算污染物在大气中的浓度,为此必须研究污染物质的各种扩散过程. 排放在大气中的污染物的扩散情况比较复杂. 影响因素比较多,这里我们仅讨论一种特殊情况.

排放在大气中的物质随风输送称为层流或平流扩散. 风的无规律运动称为湍流,由此引起的污染物的扩散称为湍流扩散. 假定污染物的扩散服从扩散定律,即单位时间通过单位法向面积的流量与它的浓度梯度成正比. 下面我们将讨论大气污染的污染物随风输送和扩散情况,即层流和湍流扩散模型.

在层流扩散中,设 Q_1 为通过与风向垂直的单位截面的层流量,U 为风的平均风速,u 为大气中的污染物浓度,显然有

$$q_1 = uU \tag{1.74}$$

在湍流扩散中,使 Q_2 为 x 方向上在单位时间内通过单位面积的扩散量,由假设它与污染物浓度的梯度成正比,有

$$q_2 = -k_x \frac{\partial u}{\partial x} \tag{1.75}$$

其中比例系数 k_x 为 x 轴方向上的湍流扩散系数,负号表示 q_2 与 $\frac{\partial u}{\partial x}$ 的方向相反.

下面来推导污染物扩散过程的数学模型. 考察空间的微小长方体元素,如图 1-11. 我们先讨论沿 x 轴方向上流量的变化. 设在单位时间内通过垂直 x 轴的单位截面污染物质的流量为 q_x,且

$$q_x = q_1 + q_2 \tag{1.76}$$

显然截面 $ABCD$ 的面积为 $dydz$,通过截面 $ABCD$ 流入该微小长方体元素的流量为 $q_x dydz$,在单位时间内流入沿 x 方向流入该微小体积元素的流量与流出量之差为

$$q_x dydz - q_{x+dx} dydz = -\frac{\partial q_x}{\partial x} dxdydz \tag{1.77}$$

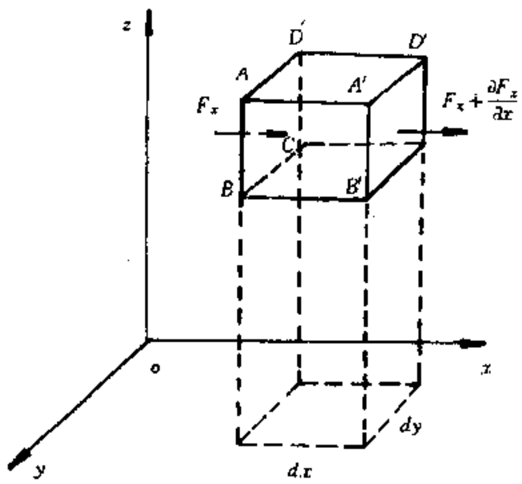


图 1-11

其中 q_{x+dx} 为从截面 $A'B'C'D'$ 的流出量,即等于$\left(q_x + \frac{\partial q_x}{\partial x}dx\right)dydz$. 将(1.74)、(1.75)和(1.76)式代入(1.77)式可得沿 x 轴方向流入量与流出量之差为

$$\left[-\frac{\partial u}{\partial x}U+\frac{\partial}{\partial x}\left(k_x\frac{\partial u}{\partial x}\right)\right]\mathrm{d}x\mathrm{d}y\mathrm{d}z \tag{1.78}$$

同理沿 y 轴方向，z 轴方向流入量与流出量之差分别为

$$\left[-\frac{\partial u}{\partial y}V+\frac{\partial}{\partial y}\left(k_y\frac{\partial u}{\partial y}\right)\right]\mathrm{d}x\mathrm{d}y\mathrm{d}z \tag{1.79}$$

$$\left[-\frac{\partial u}{\partial z}W+\frac{\partial}{\partial z}\left(k_z\frac{\partial u}{\partial z}\right)\right]\mathrm{d}x\mathrm{d}y\mathrm{d}z \tag{1.80}$$

其中 U、V、W 为平均风速在 x 轴、y 轴、z 轴方向上的分量，k_y，k_z 分别为 y 轴、z 轴方向上的湍流扩散系数.

因为流入该微小体积元素的数量与流出量之差等于该微小体积中污染物含量对时间的变化率，则有

$$\frac{\partial u}{\partial t}+\frac{\partial u}{\partial x}U+\frac{\partial u}{\partial y}V+\frac{\partial u}{\partial z}W=\frac{\partial}{\partial x}\left(k_x\frac{\partial u}{\partial x}\right)+\frac{\partial}{\partial y}\left(k_y\frac{\partial u}{\partial y}\right)+\frac{\partial}{\partial z}\left(k_z\frac{\partial u}{\partial z}\right) \tag{1.81}$$

这就是大气中污染物扩散的数量模型. 为 k_x，k_y，k_z 为相等的常数 k，则(1.81)写成

$$\frac{\partial u}{\partial t}+\frac{\partial u}{\partial x}U+\frac{\partial u}{\partial y}V+\frac{\partial u}{\partial z}W=k\left(\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}+\frac{\partial^2 u}{\partial z^2}\right) \tag{1.82}$$

若 $v=U\vec{i}+V\vec{j}+W\vec{k}$，$\nabla^2 u=\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}+\frac{\partial^2 u}{\partial z^2}$，$gradu=\frac{\partial u}{\partial x}\vec{i}+\frac{\partial u}{\partial y}\vec{j}+\frac{\partial u}{\partial z}\vec{k}$，则(1.82)式简写成

$$\frac{\partial u}{\partial t}+v\cdot gradu=k\nabla^2 u$$

习题一

1. 在一个巴基斯坦洞穴里，偶然发现了具有古代尼安德特人特征的人骨碎片，科学家们把它们带到实验室，作 C^{14} 年代测定. 分析表明 C^{14} 与 C^{12} 的比例仅是活组织的6.24%，问此人生活在多少年以前.（假定生物体死亡后，放射性碳便以每年八千分之一的速度减少）

2. 对于《洗足》、《读乐谱的妇人》这两幅画，利用表 1-1 中的数据，计算每克普通铅中白铅最初每分钟衰变的量，从而证明两幅画都是伪造 Vermeer 的.

3. 某医学专家给出了某人身上的一个肿瘤生长的一组数据如下表：

时　间 t	1	2	3	4	5	6	7
肿瘤体积	0.16	1.50	3.61	14.24	22.59	26.47	37.00

试选择适当模型，用 Mathemetica 软件求出此肿瘤生长曲线.

4. 试讨论放射疗法对肿瘤生长的影响.

5. 若在两个减肥模型中，取 $a=0.13$，$b=0.001\ 6$，试比较模型解曲线有什么差异.

6. 人体在平衡状态下的平衡值有什么实际意义.

7. 本章的三个销售模型各有哪些优缺点？

8. S-M 模型是否有解析解？若有请给出具体表达式；若没有，请给出你的对策.

9. 在模型$\frac{dx}{dt}=-by$，$\frac{dy}{dt}=-cx$中，令$x_0=15\ 000$，$y_0=20\ 000$，$b=0.04$，$c=0.5$，假定y部队有5 000可用作增援的部队，但这些部队必须是在需要的时突然地投入战斗. 对增援出现的最佳点给出一个全面的分析.

10. 设线性律模型为$\frac{dx}{dt}=-gxy$，$\frac{dy}{dt}=-hxy$，而$g=0.04$，$h=0.05$时，给出与9题类似的分析.

11. 证明$\frac{dN}{dt}=-\alpha N\ln\frac{N}{K}$的正平衡点$N=K$是全局稳定的. 这里$\alpha$和$K$为正的常数.

12. 有一种鼠繁殖得很快，如果把时间取成月，它的纯增长率是每月40%，下表给出了观测数据. 试选用适当微分方程模型. 并且评价所建立的模型与观测值的拟合情况.

月	0	2	6	10
观测值	2	5	20	109

13. 推导Greenshields模型的方法，类似可得Greenshields模型.

$$q=v_{max}\ln\frac{\rho_{max}}{\rho}$$

求流量峰值处的密度值为多少？并计算最大流量.

14. 利用一维扩散模型类似的方法推出三维扩散模型.

15. 炮弹在空中爆炸后，放出的烟雾以爆炸点为中心向四周迅速扩散. 试建立烟雾的扩散方程.

第二章　离散数学模型

图论是广泛应用在物理学、化学、控制论、信息论、科学管理、电力通讯、编码理论、可靠性理论、电子计算机等各个领域的离散数学重要分支. 在生产、科学实践中，图论的理论和方法，可以提供给我们很多简便可行的解决问题的方法. 例如：在组织生产中，各工序怎样衔接，才能使生产任务完成的既快又好. 一个邮递员送信，要走完他所负责的全部街道，完成任务后回到邮局，应该按照怎样的线路走，所走的路程最短？再如：各种通信网络的合理架设，交通网络的合理分布等，都可以用图论的方法求解.

图论历史源远流长，远在 18 世纪(1736 年)欧拉发表了一篇论文，解决了著名的哥尼斯堡(konisberg)七桥问题.

当时，哥尼斯堡城有一条普雷格尔河(pregel)，河中有两个小岛，河两岸和河中两岛，通过七座桥彼此相连. 如图 2-1

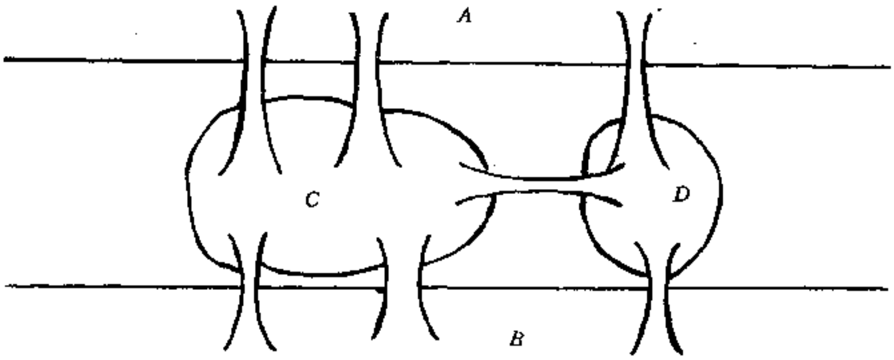


图 2-1

有一趣味问题，游人从两岸 A,B 或两个小岛 C,D 中任一个地点出发，要找到一条路线做到每座桥恰通过一次而最后返回原地. 成千上万的人在解决这个问题时都失败了，Euler 的文章给出了解决问题的答案. 他指出七桥问题是无解的，他将问题归结为如图"一笔画"问题：即能否从某一点开始，不重复地一笔画完整个图形. 如图 2-2

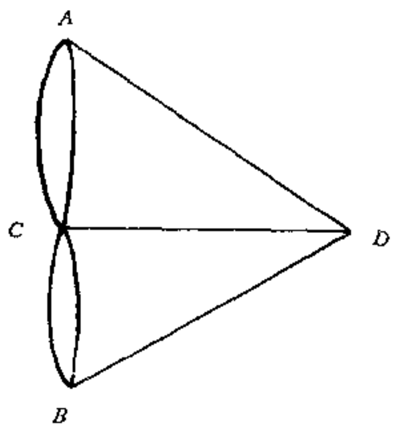


图 2-2

欧拉回答的根据是：若图是连通的，且图中奇点(和奇数条边相关联的点)的数目为 0 或 2 时，则图能"一笔画". 奇点的数目为零时，则图中任一点即是"一笔画"的起点又是终点. 奇点的数目为 2 时，则两

点中任一点为"一笔画"的起点，而另一点为终点. 图 2-2 中奇点数目为 4，不满足"一笔画"规则，因此是无解的.

哥尼斯堡七桥问题，用图论的方法得到了简捷的证明，所以欧拉被人们公认为图论的创始人，人们称他为"图论之父".

近年来，图论受计算机科学蓬勃发展的刺激，发展极其迅速. 应用范围不断拓广，已渗透到诸如语言学、逻辑学、物理学、化学等及数学的其它分支中. 特别是在计算机科学中，如形式语言、数据结构、分布式系统、操作系统等方面均扮演着重要的角色. 图论与相关的网络分析也成为经常采用的方法之一.

第一节　图的基本概念

一、图、子图

在火车站的售票处或候车室里，我们常可以看到如图 2-3 的铁路交通示意图，上面的点表示车站，连接线段表示铁路，从这张图上，我们可以了解到铁路连接各个车站的情况，至于各点站之间的相对位置、铁路线形状和长短是否符合地图是无关紧要的，而它们的关系是人们所感兴趣的，由此产生了图的概念.

图的实质：顶点代表事物，边表示事物之间关系.

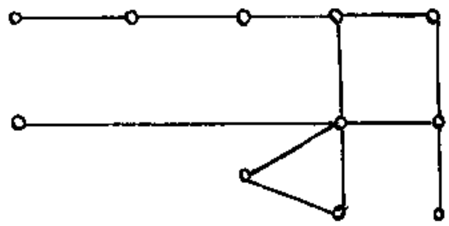


图 2-3

一个图 G 是指一个有序三元数组 (V,E,φ)，V 是非空的顶点集，E 是边集，φ 是关联函数，是连接图 G 的边和无序顶点对(不必相异)的桥梁.

若 e 是图 G 的一条边，V_i、V_j 是 $\varphi(e)=(V_i,V_j)$ 的顶点，称边 e 连接 V_i 和 V_j，顶点 V_i 和 V_j 称为 e 的端点，一条边的端点称之为与这条边关联，反之，称这条边与顶点关联. 与同一条边关联的两个顶点称为相邻的. 同理和同一个顶点关联的两条边也称为相邻的. 顶点不与任一条边关联，这顶点称孤立点，仅包含孤立点的图称为零图，记 $(n,0)$. 端点重合为一点的边称为环(或自回路). 两条或两条以上的边与同一对端点关联，称这些边为重边(或平行边)，既没有环又没有重边的图称为简单图，图 G 的顶点记为 $n(G)$ 或 $|V|$，边记为 $m(G)$ 或 $|E|$.

例 1　设 $G=(V,E,\varphi)$，其中 $V=\{v_1,v_2,v_3\}$，$E=\{e_1,e_2,e_3,e_4,e_5\}$，而 φ 定义为 $\varphi(e_1)=(v_1,v_1)$，$\varphi(e_2)=(v_1,v_2)$，$\varphi(e_3)=(v_2,v_3)$，$\varphi(e_4)=(v_2,v_3)$，$\varphi(e_5)=(v_3,v_1)$ 其图形如图 2-4.

图 2-4 中，e_1 是环，e_3、e_4 是重边，v_1 与 e_1、e_2、e_5 相关联，v_1 与 v_2，v_1 与 v_3 相邻，e_2 和 e_5 相邻. 因为图 2-4 有环，有重边，所以该图不是简单图.

例 2　设有图 $G=(V,E,\varphi)$，$\varphi(e_1)=(v_1,v_2)$，$\varphi(e_2)=(v_2,v_3)$，$\varphi(e_3)=(v_3,v_3)$，$\varphi(e_4)=(v_3,v_4)$，$\varphi(e_5)=(v_2,v_4)$，$\varphi(e_6)=(v_4,v_5)$，$\varphi(e_7)=(v_2,v_5)$，$\varphi(e_8)=(v_2,v_5)$，$\varphi(e_9)=(v_5,v_1)$，如图 2-5

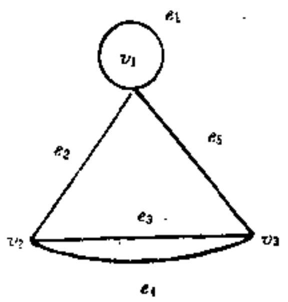


图 2-4

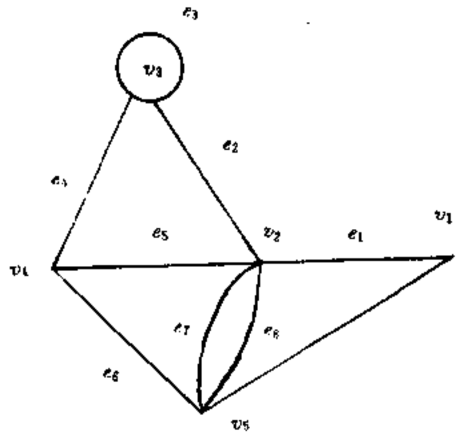


图 2-5

一个图的画法不是唯一的，只要边和顶点的关联关系不变即可，但要注意图的美观.

若一个图，它的边仅在端点处相交该图为平面图(所有顶点可以画在同一平面上)，例 2-1 是一平面图，可以用非常简便的方法在平面上表示出来.

若图只有一个顶点，该图称为平凡图，记为(1,0).

若图中任意两点间恰好有一条边，称该图为完全图，其顶点数称为图的阶，n 阶完全图记作 k_n，如图 2-6，是一个五阶完全图.

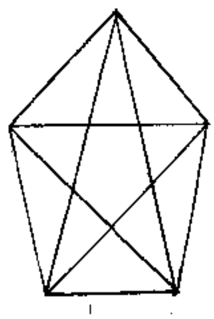


图 2-6

二部图(也叫偶图)，是指顶点可以分成两个非空子集 X 和 Y，任一条边都有一个端点在 X 中，另一个端点在 Y 中，这称为图的一个二分类. 完全二部图是具有二分类(X,Y)的简单二部图.

例 3　在遥远的地方有一位酋长，他想把三个女儿嫁出去. 假定已有 3 位求婚者，他们必须向酋长交纳一定数量的彩礼. 按当地习惯，彩礼是按牛的头数计算的. 分别记三个女儿为 A、B、C，三位求婚者为 x，y，z，每位求婚者对 A、B、C 愿出的彩礼数，视其对她们的喜欢程度而定. 现在酋长遇到的问题是，他应当怎样嫁出她的女儿，才能获得最多头数的牛？

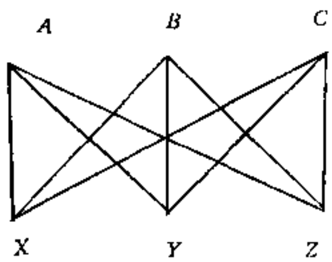


图 2-7

若令 $X=(A,B,C)$，$Y=(x,y,z)$例 2-3 是二部图的典型例题.

若图 $G=(V,E)$，$G_1=(V_1,E_1)$都是图，且 $V_1\subseteq V$，$E_1\subseteq E$，则称 G_1 是 G 的一个子图，记为 $G_1\subseteq G$.

若 $V_1\subset V$，$E_1\subset E$，则 G_1 是 G 的真子图.

若 $G_1=(V_1,E_1)$ 是 $G=(V,E)$ 的子图，且 $V_1=V$ 则称 G_1 是 G 的一个生成子图.

若 $G=(V,E)$，$V_1\subseteq V$，E_1 为 G 中两个端点都在 V_1 中的边的全部体组成的边集，则称 $G_1=(V_1,E_1)$ 为 G 关于顶点 V_1 的导出子图.（即 V' 是 V 的非空子集对 $\forall\ v_1,v_2\in V'$ 如 $(v_1,v_2)\in E$，则必有 $(v_1,v_2(\in E')$.

例 4　指出下列图中子图，真子图，生成子图，导出子图.

答：(b)，(c)，(d)是(a)的子图

(b)，(d)是(a)的真子图

(c)是(a)的生成子图

(d)是(a)的导出子图

若图 G 的所有顶点和能使图 G 成为完全图的添加边所构成的图称为 G 的补图. 记作 $\overline{G}$

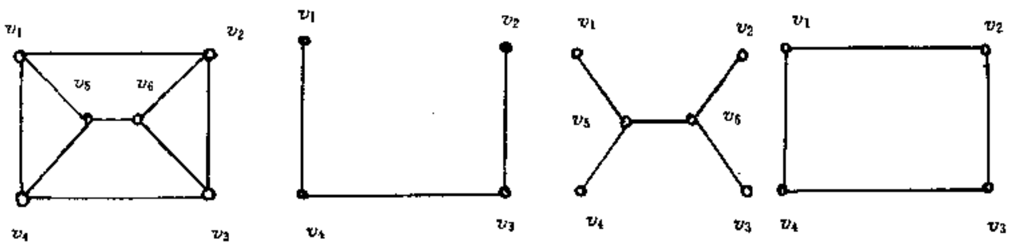


图 2-8

例 5　指出下列各图的补图 $\overline{G}$

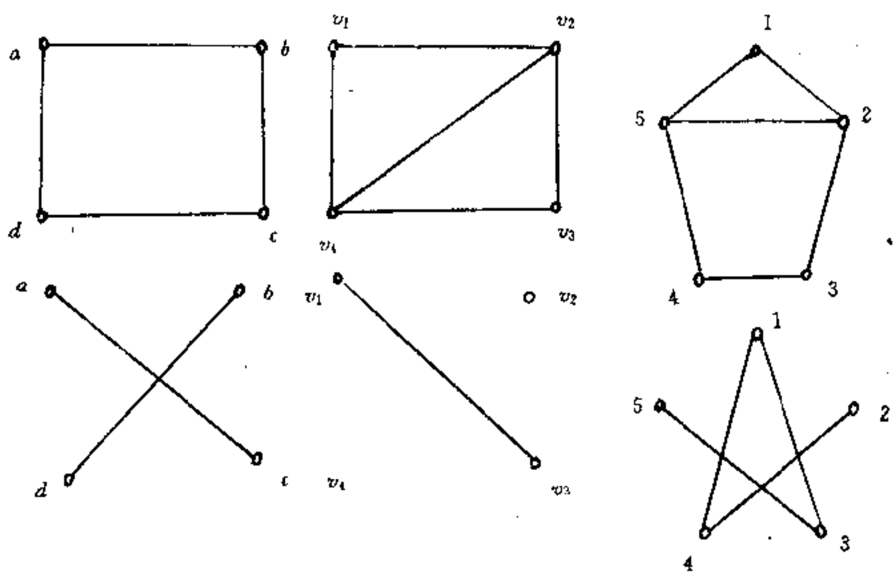


图 2-9

二、关联矩阵，邻接矩阵及顶点的度

在前面曾讨论过的图解法有其直观性，但也有局限性．当图中的结点和边的数目较大时，使用图解法就很困难，使用矩阵来表示图，能够用矩阵将图存储在计算机中，还能通过矩阵代数中的各种运算，给出图的其它性质．

关联矩阵：对任意图 $G(V,E)$ 定义一个 $n\times m$ 阶矩阵　$M=(m_{ij})_{n\times m}$（n 为顶点数，m 为边数）其中 m_{ij} 是 v_i 和 e_j 相关联的次数，该矩阵称为图 G 的关联矩阵．

邻接矩阵：$A=(a_{ij})_{n\times n}$，a_{ij} 是连接 v_i 和 v_j 的边的数目．

例 6　写出图 2-4 的关联矩阵，邻接矩阵

$$M=\begin{pmatrix}2&1&0&0&1\\0&1&1&1&0\\0&0&1&1&1\end{pmatrix}\begin{matrix}v_1\\v_2\\v_3\end{matrix}\qquad A=\begin{pmatrix}1&1&1\\1&0&2\\1&2&0\end{pmatrix}\begin{matrix}v_1\\v_2\\v_3\end{matrix}$$

$$\begin{matrix}e_1&e_2&e_3&e_4&e_5\end{matrix}\qquad\qquad\begin{matrix}v_1&v_2&v_3\end{matrix}$$

关联矩阵：邻接矩阵是用数据表示图的两种矩阵形式，由于邻接矩阵一般小于关联矩阵，因此，图在计算机中通常以邻接矩阵的形式存贮．

与一个顶点 v_i 关联的边的数目称为 v_i 的度（或次），记 $d(v_i)$，度为 1 的点为悬挂点．与悬挂点关联的边称为悬挂边，孤立点的度为零．环与顶点关联的度为 2．

例 7　写出下图的关联矩阵，邻接矩阵．

$$M=\begin{pmatrix}1&0&0&1&0&0&0\\1&1&0&0&1&0&1\\0&1&1&0&1&0&1\\0&0&1&1&0&1&0\end{pmatrix}\begin{matrix}v_1\\v_2\\v_3\\v_4\end{matrix}$$

$$\begin{matrix}e_1&e_2&e_3&e_4&e_5&e_6&e_7\end{matrix}$$

$$A=\begin{pmatrix}0&1&1&1\\1&0&1&1\\1&1&0&1\\1&1&1&0\end{pmatrix}\begin{matrix}v_1\\v_2\\v_3\\v_4\end{matrix}$$

$$\begin{matrix}v_1&v_2&v_3&v_4\end{matrix}$$

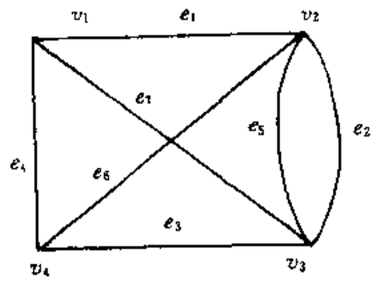


图 2-10

三、路，圈和连通

在图 $G=(V,E)$ 中一个点、边交替出现的序列 $W=v_0e_1v_1e_2v_2\cdots e_kv_k$ 称为从 v_0 到 v_k 的链，e_i 的端点是 v_{i-1} 和 v_i，v_0 和 v_k 为 W 的起点和终点，$v_1,v_2,\cdots,v_{k-1}$ 称为 W 的内部顶点．

若 W 中 $v_0,v_1,\cdots,v_k$ 互异，则称 W 为路．在路中如果 $v_0=v_k$，是称其为一个圈，记 $v_0,v_1,\cdots v_{k-1},v_0$．

在图 G 中，如果存在一条路连结顶点 u 和 v，称 u 和 v 是连通的．若图中任意两点都连通，则称 G 是连通图，否则 G 是不连通的．

最小连通图：如果从图 G 中去掉任何一条边都能使图分离，则称 G 是最小连通图．

设 S 是图 G 的一个边集，在连通图 G 中去掉 S 的全部边后，使图 G 分离为两个部分，S 的任一真子集无此性质，则边集 S 为 G 的一个割集，只含一条边的割集称为桥．割集是

使连通图失去连通性的最小的边集合.

例 8 如图 2-11(a)是一个连通图,则 $k_1=\{e_2,e_3,e_4,e_6\}$是一个割集,因为去掉这些边将分成两个连通子图如(b),$\{e_1,e_2\}$不是割集,因为去掉 e_1,e_2 图仍连通.$\{e_1,e_2,e_3,e_4\}$也不是割集,虽然去掉这些边能将图分为两部分,但不是最少的边的集合.

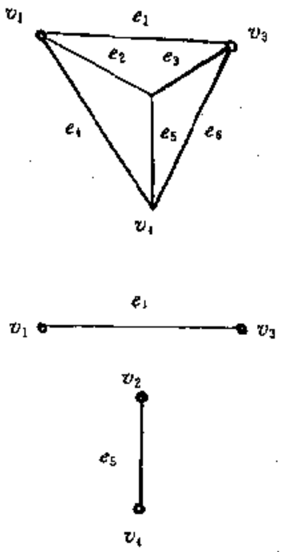


图 2-11

四、有向图,赋权图

在用图论方法解决实际问题时,仅用连接两点的边表示两者之间的关系是不够的,例如,为了表示网络上的电流方向,比赛的胜负时,仅用边表示事物关系就不确切了.因此,我们在边集上定义方向,有方向的边称为**弧**,由弧组成的图称为**有向图**,表示成 $D=(V,A,\varphi)$,V 是非空的顶点集,A 是弧集,而 φ 是关联函数,若 $\varphi(a)=(u,V)$,则称 a 的头是 V,a 的尾是 u. 在有向图 D 中,去掉边上所有的方向得到的图称为 D 的**基础图**.

在有向图中,对于任何 $v\in V$ 来说,以结点 v 为尾的边的条数,称为结点 v 的**引出次数**或**出度**,记 $d_D^+(v)$,以 v 为头的边的条数,称为结点 v 的引入数或入度,记 $d_D^-(v)$. 结点 v 的引入次数与引出次数之和,称为结点 v 的次数或**度数**.

定理 1 任意一个图(n,m)结点的度数总和等于边数的 2 倍,即 $\sum\limits_{v\in V} d(v)=2m$.

证明:∵任一条边与两个结点相关联,出现一条边使结点总度数增加 2,∴边数的 2 倍就是结点的总度数.

定理 2 n 阶完全图 k_n 的边数 $m=\dfrac{1}{2}n(n-1)$

证明:n 阶完全图,每个结点度数为 $n-1$,故结点度数和 $\sum\limits_{v_i\in V} d(v_i)=2m=n(n-1)$

$\therefore m=\dfrac{1}{2}n(n-1)$

定理 3 任意一个图,奇点个数为偶数.

证明:设 v_1,v_2 分别是图中度数为奇数和偶数结点集合,$\therefore v_1\cup v_2=V$

$\sum\limits_{v\in V_1} d(v)+\sum\limits_{v\in V_2}=\sum\limits_{v\in V} d(v)=2m$ ∵ $\sum\limits_{v\in V_2} d(v)$ 与 $\sum\limits_{v\in V} d(v)$ 是偶数 ∴ $\sum\limits_{v\in V_1} d(v)$ 必为偶数,即 v_1 奇点数为偶数.

如果在图的每一对结点之间,都有且仅有一对方向相反的有向边,则称为**有向完全图**.

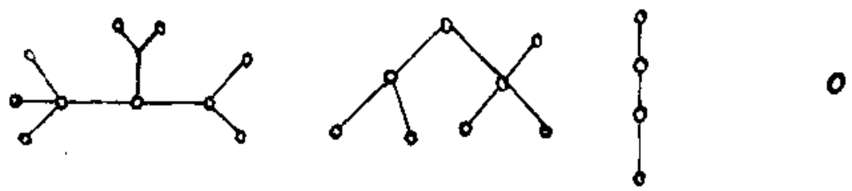
在图中,无论是在边或弧上,定义一个数 W_{ij} 与之对应,则称这样的图为**赋权图**. W_{ij} 称为边(或弧)上的**权**,权表示的是与边(或弧)相关的数量指标,如距离、时间、费用等.

第二节　最小生成树及模型

树是图论中最重要的概念之一，它是基尔霍夫在解决电路理论中求解联立方程时首先提出的. 它又是图论中结构最简单，用途最广泛的一种连通图，我们讨论树的概念，性质，主要是某些应用.

一、树的定义

定义：一个无圈的连通图称为树，如图 2-12 所示无向树.



(a)　　(b)　　(c)　　(d)

图 2-12

常用符号 $T(n,m)$ 表示一棵树，其中 n 表示树的结点数，m 表示树的边数(树枝数). 次数为 1 的结点称为树的树叶，次数大于 1 的结点称为树的分支点，边称为树枝，只有一个结点的树称为平凡树，如图 2-12(d)就是一颗平凡树.

根据树的定义，得如下结论.

定理 4：具有 n 个结点，m 条边的无向图 T 是树，当且仅当下列条件之一成立.

(1)T 无圈，且 $m=n-1$；

(2)T 连通，且 $m=n-1$；

(3)T 中任意两个顶点之间恰有一条链；

(4)T 是连通的，且每一条边都是一个割集；

(5)T 是一个最小连通图.

证明：(1)必要性：设 T 是树，根据定义 T 无圈，对于 $m=n-1$，我们用归纳法证，当 $n=1$ 时 $m=0$ 公式成立. 假设 $n=k$ 时 $m=k-1$ 成立，往证 $n=k+1$ 时，$m=k$ 成立，当 $n=k+1$ 时，即增加了一个结点，设为 v，若 v 不与任何结点连接，则 v 是孤立点，与树的定义不符，若 v 与 k 中结点连接假设 v_1，v_2，这样 v，v_1，v_2 形成圈，与树定义矛盾，因此，增加一个结点，一定增加也只能增加一条边，故此时边数 $(k-1)+1=K$，$\therefore m=n-1$ 成立.

充分性：设 T 无圈，且 $m=n-1$，往证 T 是树，(只要证明 T 连通即可). 假设 T 不连通，而分成 k 个连通分图 $T_1,T_2,\cdots,T_k$，分图结点数 n_i，边数 $m_i(1\leqslant i\leqslant k)$ $n=\sum_{i=1}^{k}n_i$，$m=\sum_{i=1}^{k}m_i$ 因为每一分图连通且无圈，每一分图都是树.$m_i=n_i-1$，$m=\sum_{i=1}^{k}m_i=n-k$ 当 $k\geqslant 2$ 时，$m=n-k<n-1$ 与题设矛盾，只有 $k=1$，T 是连通的，证毕.

(2)必要性:设 T 是树,则由定义可知 T 是连通的,且由(1)知 $m=n-1T$ 无圈.

充分性:T 连通,且 $m=n-1$,往证 T 是树,(只须证 T 无圈).用归纳法,当 $n=1$ 或 2 时,图中无圈.设 $n=k$ 时无圈,此时 $m'=n-1=k-1$ 考察 $n=k+1$ 时,增加了一个结点,设为 v,此时边的数目为 $m=k$,比 m'之了一条边,此边一端是 v,另一端与 k 个结点中某一个相关联,由于 k 个结点无圈,增加的这条边也不可能通过结点构成圈,按定义 T 是连通的,且无圈,∴T 是树.

(3)必要性:设 T 是树,则图是连通的,任意两点之间必存在通路,且只有一条,否则通路中至少有一个结点出现超过一次,因而必有圈,∴矛盾.

充分性:设(3)条件成立,则 T 是连通的,且无圈(否则通路不止一条),∴T 是树.

(4)必要性:设 T 是树,则图是连通的.如果去掉任一条边(v_i,v_j)后图仍连通,说明 v_i,v_j 之间还存在另一条通路,(v_i,v_j)构成圈,矛盾,所以去掉任一条边图不再连通,任一边为割集.

充分性:图连通,去掉任一边图不再连通,说明图中无圈,∴T 是树.

(5)必要性:设 T 是树,由(4)知,去掉任一边图不再连通,∴T 是最小连通图.

充分性:T 是最小连通图,说明去掉任一条边图不再连通,T 无圈,∴T 是树.

若图 $T=(V,E',\varphi)$是图 $G=(V,E,\varphi)$的生成子图,且 T 是一个树,则 T 是 G 的一个**生成树**.

以上结论给我们提供了求生成树的方法.

"**破圈法**":任取一个圈,从圈中去掉一边,对余下的圈重复这个步骤,直到不含圈时为止.

例 1 在图 2-13 中,用破圈法求出图的一个生成树.

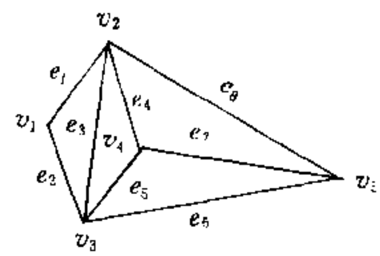


图 2-13

解:

在(v_1,v_2,v_3,v_1)中去掉 $e_3=[v_2,v_3]$

在(v_1,v_2,v_4,v_3,v_1)中去掉 $e_4=[v_2,v_4]$

在$(v_1,v_2,v_5,v_4,v_3,v_1)$中去掉 $e_8=[v_2,v_5]$

在(v_3,v_4,v_5,v_3)中去掉 $e_6=(v_3,v_5)$,这时图中不含圈,于是得一生成树.

"**避圈法**":在图中任取一条边 e_1,找一条与 e_1 不构成圈的边 e_2,再找一条与$\{e_1,e_2\}$不构成圈的边 e_3,如此重复下去,直至不能再添加任何边为止.

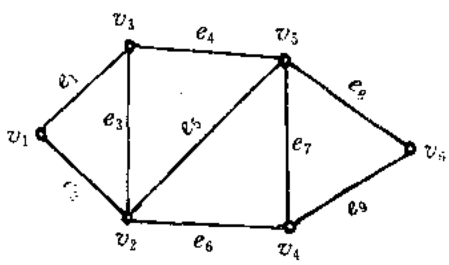


图 2-14

例 2 如图 2-14 用避圈法求一生成树.

解:先取$\{e_1,e_2\}$,再取 e_5,(不能取 e_3,否则构成圈)取 e_6(不能取 e_4),再取 e_8 $\{e_1,e_2,e_5,e_6,e_8\}$构成生成树,生成树不是唯一的.

二、最小生成树

设图 $G=(V,E,\varphi)$ 是一个非负赋权连通图，即任一边权 $W_{ij}\geqslant 0$；$T=(V,E',\varphi)$ 是 G 的一个生成树，称生成树中所有边的权之和为生成树 T 的权，记 $W(T)$，$W(T)=\sum W_{ij}$，$(v_i,v_j)\in T$，若生成树 T^* 的权 $W(T^*)$ 是 G 中所有生成树的最小者，称 T^* 是 G 的最小生成树，简称最小树．最小树不唯一，最小树的权唯一．

例 3　某工厂内联结六个车间的道路网如图 2-15 所示．已知每条道路的长，要求沿道路架设联结六个车间的电话线网，使电话线的总长最小．

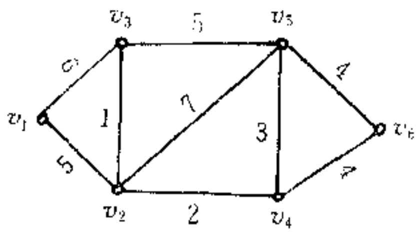


图 2-15

解：此题求赋权图上的最小树．

方法一：(避圈法 kruskal)开始选一条最小权的边 $e_1=(v_2,v_3)$，以后每一步都从未被选取的边中选一条权最小的边，并使之与已选取的边不构成圈，(每一步中，如果有两条或两条以上的边都是权最小的边，则从中任选一条)，选出 $e_1(v_2,v_3)$ 后，取 $e_2=(v_2,v_4)$，$e_3=(v_4,v_5)$，$e_4=(v_5,v_6)$[或 (v_4,v_6)]．$e_5=(v_1,v_2)$，虽然 (v_4,v_6) 的权小于 (v_1,v_2)，但因其构成圈，所以选 (v_1,v_2)，这时任一条未选的边都与已选的边构成圈，所以算法终止，电话线总长 1+2+3+4+5=15(单位)．

方法二：(破圈法)，任取一个圈，从圈中去掉一条权最大的边，(如果权最大的边有两条或两条以上，则去掉其中任意一条)，在余下的图中重复该步骤，一直到不含圈的图为止，这时得到的图便是最小树．

例 4　用破圈法求图 2-15 的最小生成树．

解：任取圈 (v_1,v_2,v_3,v_1)，将 $[v_1,v_3]$ 去掉；在 (v_2,v_3,v_5,v_2)，中将 $[v_2,v_5]$ 去掉，在 (v_2,v_3,v_5,v_4,v_2) $[v_3,v_5]$ 去掉，在 (v_4,v_5,v_6,v_4) 中去掉 (v_5,v_6) 或 (v_4,v_6)，此时得到一个不含圈的图，即为最小树．

例 5　有 8 个系或部门(SU)其中 A 为学校行政机关，SU 需要设计一个网络系统，使得从 A 到每个系都有电缆相连．SU 主管部门希望连接电缆的用量尽可能地少．施工单位测算了各个部门之产所需电缆的长度(单位：米)，如表 2-1

表 2-1

	A v_1	B v_2	C v_3	D v_4	E v_5	F v_6	G v_7	H v_8
A　v_1		80	90	120	140	120	170	220
B　v_2			70	90	110	70	120	170
C　v_3				50	170	100	150	180
D　v_4					90	100	150	150
E　v_5						80	80	160
F　v_6							60	110
G　v_7								100
H　v_8								

解法一：(kruskal)方法：先找出距离最短的电缆(v_3,v_4)，其长度为50；然后在剩余的电缆中继续寻求最短者，(v_6,v_7)入选，长度为60；继续下去，(v_2,v_3)，(v_2,v_6)的长度均为70，任取一条(v_2,v_3)，接下去(v_2,v_6)也是解的一部分；(v_1,v_2)，(v_5,v_6)入选，(v_5,v_7)长度与(v_1,v_2)，(v_5,v_6)相同，但若连接v_5,v_7则会产生一个圈，所以避开(v_5,v_7)，(v_2,v_4)，(v_4,v_5)，(v_1,v_3)是剩下的电缆中长度最短的，但连接它们会产生圈，所以再选，(v_3,v_6)，(v_4,v_6)长度最短，但连接它们会产生圈，所以选(v_7,v_8)，此时已有七条电缆，网络最小生成树完成．见图2-16，其电缆总长度(总权)为：50＋60＋70＋70＋80＋80＋100＝510米．注意：由于出现相同权的边，所以最小生成树不是唯一的．

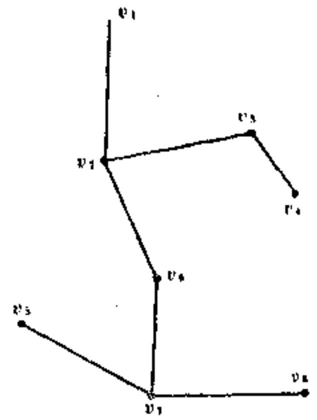


图 2-16

解法二：Prim 方法：在图2-16中任选一系v_1，作为解的一部分，称其为一个分枝，在剩下七个系中找出距v_1最近的系v_2，其长度为80，(v_1,v_2)为解的一部分，连接v_1,v_2，在剩下的系中找出距v_1,v_2最近的系v_3,v_6，长度各为70，此时该分支(v_1,v_2,v_3,v_6)继续下去，下一个系v_4距$v_3$50米，v_6距$v_7$60米，v_6距$v_5$80米；v_7距$v_8$100米，该网络总长度80＋70＋70＋50＋60＋80＋100＝510米，最小生成树如图2-17

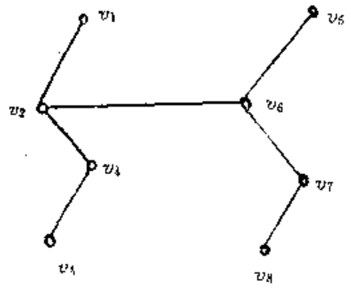


图 2-17

求最小生成树的 Prim 算法步骤

(1)在图G中任取一顶点，设其为v_1，令$k=1$，对其用零值给以永久标号$(p(v_k)=0)$；令$c=\{v_k\}$，$A=\phi$，c表示具有p标号的点的集合，A表示生成树的边集．对其它所有的顶点$v_i\in V\backslash C$给以临时标号$T(v_i)=\omega_{ik}$(ω_{ik}为(v_i,v_k)边的权，当v_i,v_k不相连时，记$\omega_{ik}=\infty$)．

(2)在所有临时标号中，选择最小权值的一个使其成为永久标号．设$T(v_j)=\min\limits_{v_i\in V/C} TV\backslash C(v_i)$，将$v_j$的$T$标号变为$p$标号$p(v_j)=T(v_j)$；置$c=c\cup\{v_j\}$，$A=A\cup\{(v_j,v_k)\}$，令$k=j$．

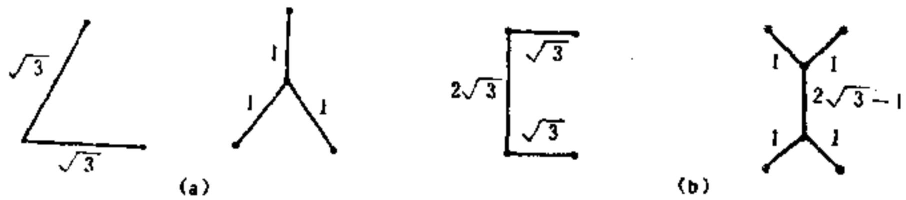
(3)临时标号不存在时，则停止；否则对每一个具有临时标号的点$v_i\in V\backslash C$，令$T(v_i)=\min\{T(v_i),\omega_{ik}\}$转第二步骤(2)．

两种方法相比：Prim 方法避免了每次增加边时必须检验的要求，也避免了比较剩下的边的权，因此 Prim 方法比 kruskal 方法更有效．

三、通讯网络的最小生成树

通讯站间通讯线路的费用与线路的长度成正比．人们经常通过引入若干个"虚设站"

并构成所谓 Steiner 树的方法来降低通讯站的最小生成树的费用. 节省率达到 $\left(1-\frac{\sqrt{3}}{2}\right)$ 即 13.4%,并且为构造有 n 个通讯站的网络的费用最低的 Steiner 树,只需 $n-2$ 个虚设站. 例如:如下两个简单例子.

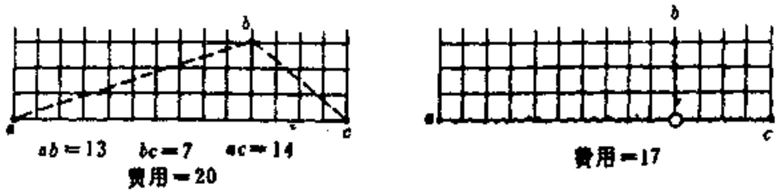


(a)　　(b)　　(c)　　(d)

图 2-18　原始最小生成树和与之对应的 Steiner 树

对于局部的通讯网络,经常用直折线(或方格距离)来代替欧氏直线距离. 计算方法如下.

例 6　设有 9 个通讯站的局部网络需要设计一个最小费用的生成树,它们的直角坐标为:$a(0,15)$,$b(5,20)$,$c(16,24)$,$d(20,20)$,$e(33,25)$,$f(23,11)$,$g(35,7)$,$h(25,0)$,$i(10,3)$



ab=13　bc=7　ac=14
费用=20　　费用=17

图 2-19

若用方格距离而且所有虚设站必须位于格子点上(即其坐标为整数),每条直线段的费用就是其长度. 在求其最小费用生成树之前先解决以下几个问题:

(1)是否存在最小费用生成树?

(2)最多要增加几个虚设站?

(3)如何确定虚设站的位置及简化?

(4)怎样找到最小费用生成树?

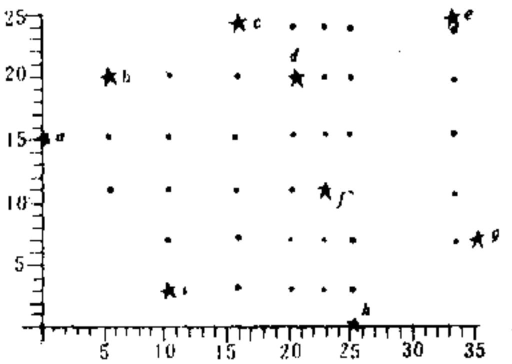


图 2-20　31 个虚设站位置　带 * 号原始数据标号

根据 Steiner 树有以下结论

(1)存在一个最小的 Steiner 树,可以横跨整个通讯网.

(2)有 n 个顶点网络最多需要 $n-2$ 个虚设站.

(3)由于虚设站必须在格子点上，由 9 个站直角坐标，$0\leqslant x\leqslant 35$，$0\leqslant y\leqslant 25$，虚设站的位置可达 $36\times 26=936$ 个，根据下面的定理，可将虚设站减少到 31 个点的位置(见图 2-20)

(4)根据已知点和虚设点可求得最小生成树.

定理 1　如果不考虑添加站的费用，对于任何一个最优化结构，所有虚设站的坐标都必须位于由 n 个基本点横坐标 x_i 和纵坐标 y_j 组合而成的点上，$(i,j=1,2,,3,\cdots,n)$.

定理 2　如果所有的基本点 (x_i,y_i) 在点 (x',y') 的上方或右方(上方或左方，下方或右方，下方或左方)，则 (x',y') 不是虚设站.

由定理 1 可知虚设站 $8\times 9-9=63$ 个(将已知 9 个通讯站去掉). 由定理 2 可知只有上、下、左、右都有基本点的 (x',y') 才能作为虚设站. 将边角上的 32 个位置去掉，最终虚设站的位置减少至 31 个. 若用穷举法：$1+C_{31}^{1}+C_{31}^{2}+\cdots+C_{31}^{7}=3\ 572\ 224$ 个情形的最小生成树，从中选出最小一个，长度为 94，但如果问题的规模增大时，其计算时间，计算量均会增加，甚对无法求解. Steiner 树属于图论中 NPC 问题(穷举法). 对于该通讯网络的虚设站的设立我们采取"模拟退火法".

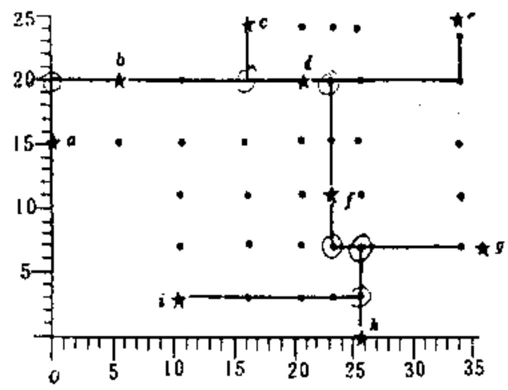


图 2-21　模拟退火法最优解之一

基本思想：将已知网络点分为三个顶点一组，通过增加虚设站的方法优化每一组，最后整体化，多次迭代，收敛到最优结果中的一个，如图 2-21

第三节　最短路模型

最短路问题

定义 1：在带权有向图 D 中，弧 (v_i,v_j) 上带的权 $w(v_i,v_j)$ 称为这条弧的长度. 一条路径的长度是这条路径所经过的弧的长度和. 即若 p 是 D 中的一条路径，则 p 的长度为 L_p 为　$L_p=\sum\limits_{(u,v)\in p}w(u,v)$

定义 2：在带权图 D 中，若 P 是从结点 u 到结点 v 的所有路径中最短的一条路径，则称 p 为从 u 到 v 的最短路径，其长度记作 $L(u,v)$ 并称为从 u 到 v 的距离.

最短路径具有重要的实际意义. 例如：用结点表示城市，弧的长度表示一个城市到另一个城市公里路程，则最短路径问题就是寻找从一城市到另一城市的最短里程问题，如果弧的长度表示修筑道路的经费，则最短的路径表示最小投资等等.

最短路径的定义可以推广到带权无向图中.

怎样寻找从指定点到其它各点的最短路径呢？1959 年得克斯特拉(Dijkstra)提出：用

逐点增长的方法构造一颗路径树，从而得到从树根(已知点)到其它点的距离，这是目前普遍采用的一个较好的算法，具体作法如下：

求从指定点 v_0 到某一点 v 的最短路径.

从 v_0 出发沿着弧的方向第一个到达而且路径最短的一点，一定是所有与 v_0 有弧关联且弧的长度最小的那个点，这是与 v_0 距离最短的一点，称为第一短路径点，如果这个点就是 v，问题得到解决，如果不是 v 而是另一点 v_1，那么从 v_0 出发第二条短路径会到达哪个点呢？一定到达与 v_0 有弧关联且弧长度最小的那个点(v_1 除外)，或者是由 v_0 经过 v_1 再经过与 v_1 关联的长度最小的一条弧所到达的点，取两条路径中最短的一条，它的终点即是从 v_0 出发第二条短路经所到达的点，设为 v_2，如果 v_2 即为所求的 v，问题即得解决，否则继续寻找从 v_0 出发第三短路径到达的点，第四短路径到达的点，如此反复下去直到路径的终点是 v 为止.

设 T 为已求得的第一、第二、…第 k 条短路径的终点集合，即 $T=\{v_1,v_2,\cdots,v_k\}$，则有 $L(v_0,v_1)\leqslant L(v_0,v_2)\leqslant\cdots\leqslant L(v_0,v_k)$. 那么第 $k+1$ 条短路径(设其终点为 w)或者是弧 (v_0,w)，或者终点是 T 外的某一点 w，但路径的中间结点一定都是 T 中的结点(即已到达过的结点)这是显然的，如果从 v_0 到 w 的路径中有一中间点 i 不在 T 中，由于 $L(v_0,i)<L(v_0,w)$ 则 i 点必然作为比 v_0 到 w 的路径更短的一条路径终点先行加入 T 中. 因此，按照路径逐步增长的过程，我们将得到一棵以 v_0 为根的树，称为路径树.

如何求路径端点？

设已求出从 v_0 出发长度依次递增的 k 条路径，它们的端点各为 $v_1,v_2,\cdots,v_k$，并归入已到达的端点的集合 T 中，将 $L(v_i)$ 表示为 v_0 到 v_i 距离，对 T 外的任一结点 v'，计算数值

$$L(v')=\min_{0\leqslant i\leqslant k}\{L(v_i)+w(v_i,v')\}$$

约定 $L(v_0)=0$，若 $(v_i,v')\notin E$，则令 $w(v_i,v')=\infty$(计算机中用某一允许的最大值代替 ∞).

因此，第 $k+1$ 条短路径，其端点一定是 T 外各点中 $L(v)$ 最小的那个点(v_0 已加入 T 中)，设点为 v_{k+1}

$$L(v_{k+1})=\min\{L(v')\mid v'\notin T\}$$

直接应用算式计算是很麻烦的，我们对上式加以改进，在求最短路径的过程中，不断修改未通过点的路径参数，得如下算法：

Dijkstra 最短路径算法：

1. for all $v\neq u$ $L(v)\leftarrow w(u,v)$
2. $L(u)\leftarrow 0$
3. $T\leftarrow\{u\}$
4. while $T\neq v$ *do*

   begin
5. 找一点 $v'\notin T$ 且对所有 $v\notin T$ $L(v')\leqslant L(v)$
6. $T\leftarrow T\cup\{v'\}$
7. for all $v\notin T$

   if $L(v)>L(v')+w(v',v)$

   then $L(v)\leftarrow L(v')+w(v',v)$

end

说明：指定结点设为 u，$w(u,u)=0$，T 为通过的结点集合，初始值为 $\{u\}$，当 $T=V$ 时表示图的所有结点已通过，计算结束. 计算从第 4 行循环语句开始，在不属于 T 的结点中选路径长度最小的一点 v'，将 v' 加入 T 中并根据 v' 修改不属于 T 的结点的路径长度，如此反复进行直到所有结点都加入 T 中为止，因此这一算法是求从指定点 u 到所有其余结点的距离，其值由通过结点 v' 的 $L(v')$ 给出. 如果只需要求 u 到某些点的距离，那么只要对第 4 行的循环语句稍作修改即可.

例 1　求图 2-22 中结点 u 到其余各点的距离.

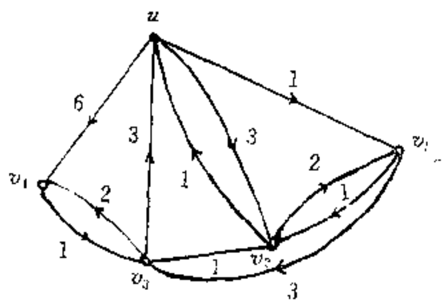


图 2-22

解：$L(v_1)=1, L(v_2)=3, L(v_3)=\infty, L(v_4)=6, L(u)=0, T=\{u\}$

(1)选择 T 外路径最小一点 v_1，于是有 $T=\{u,v_1\}$

根据 v_1 修改 T 外各点路径因 $L(v_1)+w(v_1,v_2)=1+1<L(v_2)=3$，$\therefore L(v_2)\leftarrow 2$

同理　$L(v_3)\leftarrow 1+3=4$　$L(v_4)=6$(不改变)

(2)选择 T 外路径最小一点 v_2，于是有 $T=\{u,v_1,v_2\}$根据 v_2 修改 T 外各点路径.

因　$L(v_2)+w(v_2,v_3)=2+1<L(v_3)=4$　$L(v_3)\leftarrow 3$

$L(v_4)=6$　(不改变)

(3)选择 T 外路径最小一点 v_3，于是有 $T=\{u,v_1,v_2,v_3\}$

根据 v_3 修改 T 外各点路径，$L(v_3)+w(v_3,v_4)=3+2<L(v_4)=6$

$\therefore L(v_4)\leftarrow 3+2=5$

(4)得到最后一点 v_4，于是有 $T=\{u,v_1,v_2,v_3,v_4\}$最后 $T=v$，计算结束，u 到各点距离 $L(v_1)=1$　$L(v_2)=2$　$L(v_3)=3$　$L(v_4)=5$.

若用 Dijkstra 法求指定点到其余各点的最短路径，增加一路径变量 $p(v_i)$，$p(v_i)=\begin{cases}\{u\}\cup\{v_i\} & \text{若 } w(u,v_i)=\infty \\ \phi & \text{若 } w(u,v_i)=\infty\end{cases}$ 经过各次迭代和修改过程，最后得到的 $p(v_i)$和 $L(v_i)$即为指定点 u 到点 v_i 的最短路径和距离，算法如下：

1. for all $v\in V$，给出 $L(v)$和 $p(v)$的初始值.
2. $T\leftarrow\{u\}$
3. while $T\neq V$ do

   begin

4. 找一点 $v'\notin T$，且对所有 $v\notin T$ 有 $L(v')\leqslant L(v)$
5. $T\leftarrow T\cup\{v'\}$
6. for all $v\notin T$

   if $L(v)>L(v')+w(v',v)$

   then $L(v)\leftarrow L(v')+w(v',v)$

$p(v) \leftarrow p(v') \cup \{v\}$

end

例 2 求图 2-23 中结点 v_0 到其余各点的最短路径和距离.

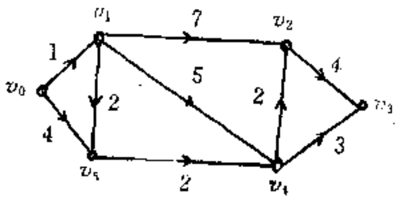


图 2-23

解:由图得出结点 $L(v)$和 $p(v)$的初始值,计算的迭代过程列成表如下,v'即为每次访问结点.

该行的 $p(v)$和 $L(v)$即为 v_0 到该点的最短路径和距离.

表 2-2

迭代次数	v_0	v_0		v_1		v_2		v_3		v_4		v_5		T
		$L(v_0)$	$p(v_0)$	$L(v_1)$	$p(v_1)$	$L(v_2)$	$p(v_2)$	$L(v_3)$	$p(v_3)$	$L(v_4)$	$p(v_4)$	$L(v_5)$	$p(v_5)$	
0	1	0	$\{v_0\}$	1	$\{v_0,v_1\}$	∞	φ	∞	φ	∞	φ	4	$\{v_0,v_5\}$	$\{v_0\}$
1	v_1			1	$\{v_0,v_1\}$	8	$\{v_0,v_1,v_2\}$	∞	φ	6	$\{v_0,v_1,v_4\}$	3	$\{v_0,v_1,v_5\}$	$\{v_0,v_1\}$
2	v_5					8	$\{v_0,v_1,v_2\}$	∞	φ	5	$\{v_0,v_1,v_5,v_4\}$			$\{v_0,v_1,v_5,v_4\}$
3	v_4					7	$\{v_0,v_1,v_5,v_4,v_2\}$	8	$\{v_0,v_1,v_5,v_4,v_3\}$	5	$\{v_0,v_1,v_5,v_4\}$			$\{v_0,v_1,v_5,v_4\}$
4	v_2					7	$\{v_0,v_1,v_5,v_4,v_2\}$	8	$\{v_0,v_1,v_5,v_4,v_3\}$					$\{v_0,v_1,v_5,v_4,v_2\}$
5	v_3							8	$\{v_0,v_1,v_5,v_4,v_3\}$					$\{v_0,v_1,v_5,v_4,v_2,v_3\}$
6						$T=V$	停止							

对于赋权无向图 $G=(V,E)$,因为沿边$[v_i,v_j]$既可,可以从 v_i 到达 v_j,也可以沿 v_j 到达 v_i,所以边(v_i,v_j)可以看作具有相同权的两条弧(v_i,v_j)及(v_j,v_i)同样可以求出从结点 u 出发到各点最短路.

例 3 如图 2-24 给出学校系与系之间距离一个人从 v_1 系出发到 7 个部门去办事,从现有的道路网络中确定他如何行走,到各系的路程最短,系之间距离用 w_{ij}表示,∞表示没有道路相连.

解:我们用 Dijkstra 方法解决校园最短路问题.用 $p(v_j)$表示从 v_i 到某点 v_j 的最短

路. 在本网络中 $v_s=v_1$, $d(v_1,v_1)=0$, v_1 具 p 标号点 $p(v_1)=0$, 与 v_1 关联的边有 (v_1,v_2) 和 (v_1,v_3), $w_{12}=80$, $w_{13}=90$, 从 v_1 出发沿 (v_1,v_2) 到 v_2 路程 $d(v_1,v_1)+w_{12}=80$. 从 v_1 出发沿 (v_1,v_3) 到 v_3 路程为 $d(v_1,v_1)+w_{13}=90$. 因为, $\min\{d(v_1,v_1)+w_{12}, d(v_1,v_1)+w_{13}\}=80$, 从 v_1 到 v_2 最短路是 (v_1,v_2), $d(v_1,v_2)=80$; 否则从其它路径走距离都比 80 大. 从而 v_2 是离 v_1 最近, v_2 成为具 p 标号点, $p(v_2)=80$; 继续考察 v_1, v_2 到其它各系(顶点)道路有 (v_1,v_3), (v_2,v_6), (v_2,v_3), (v_2,v_5), 从 v_1 出发, 分别沿最短路 (v_1,v_1) 到 v_3, (v_1,v_2) 到 v_3, (v_1,v_2) 到 v_5, (v_1,v_2) 到 v_6 的路长, $d(v_1,v_1)+w_{13}=90$, $d(v_1,v_2)+w_{23}=150$, $d(v_1,v_2)+w_{25}=190$, $d(v_1,v_2)+w_{26}=150$, 其中最短的一条为沿 (v_1,v_1) 到 v_3 的路最短, 令 $p(v_3)=90$, v_3 成为具 p 标号的点. 如此反复下去, $p(v_4)=140$, $p(v_6)=150$, $p(v_5)=190$, $p(v_7)=210$, $p(v_8)=260$. 所有顶点成为永久标号时停止运算. 表 2-4 给出从 v_1 出发到其它各系最短路径.

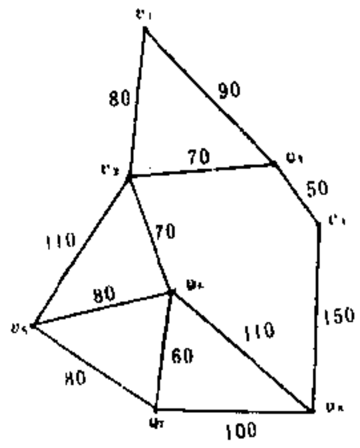


图 2-24

表 2-3

	v_2	v_3	v_4	v_5	v_6	v_7	v_8
v_1	80	90	∞	∞	∞	∞	∞
v_2		70	∞	110	70	∞	∞
v_3			50	∞	∞	∞	∞
v_4				∞	∞	∞	150
v_5					80	80	∞
v_6						60	110
v_7							100
v_8							

表 2-4　v_1 到其它各系最短路及路长

系	路	路长
v_1	v_1	$p(1)=0$
v_2	v_1-v_2	$p(2)=80$
v_3	v_1-v_3	$p(3)=90$
v_4	$v_1-v_3-v_4$	$p(4)=140$
v_5	$v_1-v_2-v_5$	$p(5)=190$
v_6	$v_2-v_2-v_6$	$p(6)=150$
v_7	$v_1-v_2-v_6-v_7$	$p(7)=210$
v_8	$v_1-v_2-v_6-v_8$	$p(8)=260$

Dijkstra 算法只适合于权值作为非负实数的情形，如果权值有正有负，该算法失效，1964 年福特提出了适用于负权边（弧）的最短路径的算法，由于篇幅所限，我们不再介绍具体做法.

第四节　网络最大流模型

在生产实践当中，许多系统都包含了流量问题. 例如：供水系统中的水流，控制系统中的信息流，公路系统中的车辆流，金融系统中的现金流等等.

例 1　图 2-25 是联结某种产品产地 v_1 和销地 v_6 的交通网，弧 (v_i,v_j) 代表从 v_i 到 v_j 的运输线，产品经弧由 v_i 输出到 v_j，弧旁的数字表示该弧所代表的运输线的最大通过能力，产品经过交通网从 v_1 输送到 v_6，求：制定一个运输方案使从 v_1 运到 v_6 的产品数量最多.

图 2-26 给出了一个运输方案，每条弧旁的数字表示在这个方案中，每条运输线上的运输数量. 这个方案可以使 8 个单位产品从 v_1 运到 v_6，该交通网上的输送量是否还可以增多，或者说该运输网络中，从 v_1 到 v_6 最大输送量是多少呢？本节就要研究类似的问题.

一、基本概念与基本定理

1. 网络与流

定义 1：给定一个有向图 $D(V,A)$，在 V 中指定了一点，称为发点（记为 v_s）和另一点，称为收点（记为 v_t），其余的点叫中间点. 对每一弧 $(v_i,v_j)\in A$ 对应有一个 $(v_i,v_j)\geqslant 0$（或简写为 C_{ij}），称为该弧的容量，一条弧的容量可以看作为沿着这条弧输送物品所允许的最大流量，通常这样的图 D 叫做一个网络，记 $D=(V,A,C)$

网络上的流指定义在 A 上的一个函数 $f=\{f(v_i,v_j)\}$，称 $f(v_i,v_j)$ 为弧 (v_i,v_j) 上的流量. 简记为 f_{ij}，图 2-25 就是一个网络，指定 v_1 是发点，v_6 是收点，其它的点是中心点，弧旁的数字为 C_{ij}.

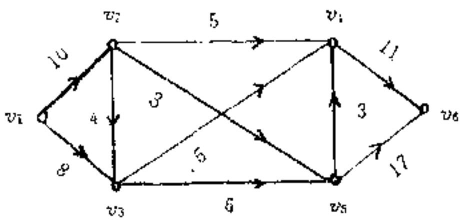


图 2-25

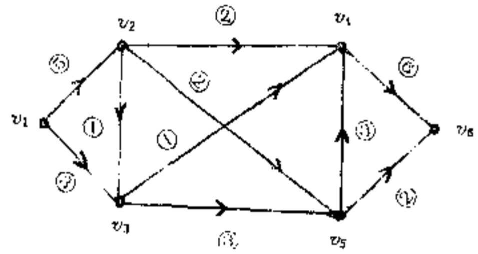


图 2-26

图 2-26 所示运输方案，可以看作这个网络上的一个流，每个弧上的运输量就是该弧上的流量，即　$f_{12}=5, f_{13}=3, f_{24}=2, f_{34}=1$ 等等.

2. 可行流与最大流

在网络问题中，对于流有两个明显要求

(一)每个弧上的流量不能超过该弧容量；

(二)中间点的流量为零.

对于每个点，运出这点的产品总量与运进这点的产品总量之差，是这点的净输出量，简称为这一点流量，中间点只起运转作用，所以中间点流量必为零. 发点的净流出量等于收点的净流入量，是方案的总输出量.

定义 2：满足下列条件的流 f 称为可行流；

(1)容量限制条件：对每一弧 $(v_i,v_j)\in A$，$0\leqslant f_{ij}\leqslant c_{ij}$

(2)平衡条件：中间点　流出量＝流入量，即对每个 $i(i\neq s,t)$ 有

$$\sum_{(v_i,v_j)\in A} f_{ij}-\sum_{(v_j,v_i)\in A} f_{ji}=0$$

对于发点 v_s，记 $\sum\limits_{(v_s,v_j)\in A} f_{sj}-\sum\limits_{(v_j,v_s)\in A} f_{js}=v(f)$

对于收点 v_t，$\sum\limits_{(v_t,v_j)\in A} f_{tj}-\sum\limits_{(v_j,v_t)\in A} f_{jt}=-v(f)$

$v(f)$称为可行流的流量，即发点的净输出量(或收点的净输入量).

可行流总是存在的，如今所有弧的流量 $f_{ij}=0$ 得到一个可行流(称为零流)，流量 $v(f)=0$

最大流问题，就是求一个流$\{f_{ij}\}$，使其流量 $v(f)$达到最大，且满足

$0\leqslant f_{ij}\leqslant c_{ij}$，$(v_i,v_j)\in A$

$$\sum f_{ij}-\sum f_{ji}=\begin{cases} v(f) & i=s \\ 0 & i\neq s,t \\ -v(f) & i=t \end{cases}$$

对于最大流问题，用线性规划的方法可以解决，但我们如果利用图的方法解决更方便，更直观.

3. 可增路

任给一个可行流 $f=\{f_{ij}\}$，我们把网络中使 $f_{ij}=c_{ij}$的弧称为饱合弧，使 $f_{ij}<c_{ij}$的弧称为非饱合弧. 把 $f_{ij}=0$ 的弧称为零流弧，$f_{ij}>0$ 的弧称为非零流弧. 图 2-26 中(v_5,v_t)是饱合弧，其它的都是非饱合弧，所有的弧称非零流弧.

定义 3：设 f 是网络的一个可行流，对于网络中每条路 p，我们在其上定义一个非负函数.

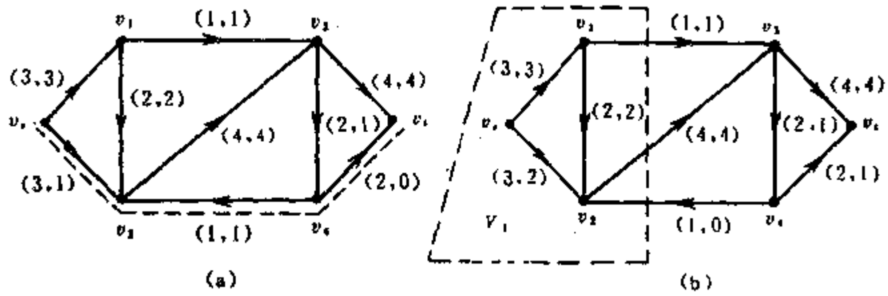
$$\delta(p)=\min_{(v_i,v_j)\in P}\begin{cases} c_{ij}-f_{ij} & 若\ a_{ij}是\ P\ 顺向弧 \\ f_{ij} & 若\ a_{ij}是\ P\ 反向弧 \end{cases}$$

若 $\delta(P)>0$ 说明沿着 P 可以达当增加流量，从发点 v_s 到收点 v_t 的 f 非饱和路称为 f 可增路.

如图 2-27，路 v_s,v_2,v_4,v_t 是一条 f 可增路，P 的顺向弧(v_s,v_2)，(v_4,v_t)，P 的反向弧(v_2,v_4)，并且 $\delta(P)=1$，f 可增路 P 的存在表示 f 不是最大流，实际上沿着 P 增加一个值为 $\delta(P)$的附加流量，得到由

$$\hat{f}_{ij}=\begin{cases} f_{ij}+\delta(P) & 若\ a_{ij}\ 是\ P\ 的顺向弧 \\ f_{ij}-\delta(P) & 若\ a_{ij}\ 是\ P\ 的反向弧 \\ f_{ij} & 其它 \end{cases}$$

所定义的新可行流 $\tilde{f}$，$v(\tilde{f})=v(f)+\delta(P)$ 称 $\tilde{f}$ 为基于 P 的调整流，图 2-28 是图 2-27 基于 f 可增路 v_s,v_2,v_4,v_t 的调整流．



弧旁的数字为(c_{ij},f_{ij})

图 2-27　一条可增路 P　　　　图 2-28　基于 P 的调整流

4．割集与割的容量

设 $S,T\subset V$，$S\cap T=\varphi$，将始点在 S，终点在 T 中的所有弧构成的集合，记为(S,T)．

定义 4：给网络 $D=(V,A,C)$，若点集 V 被剖分为两个非空集合 $V_1,\bar{V}_1$，使 $v_s\in V_1$，$v_t\in\bar{V}_1$，则把弧集$(V_1,\bar{V}_1)$称为是(分离 v_s 和 v_t)割集．

把割集弧从网络中去掉，v_s 和 v_t 之间便不存在路，所以割集是从 v_s 到 v_t 的必经之道．

定义 5：给一割集$(V_1,\bar{V}_1)$，把割集$(V_1,\bar{V}_1)$中的弧的容量之和称为这个割集的容量，记为 $C(V_1,\bar{V}_1)$，即 $C(V_1,\bar{V}_1)=\sum\limits_{(v_i,v_j)\in(V_1,\bar{V}_1)}C_{ij}$

易知：任一可行流的流量 $v(f)$ 都不会超过任一割集的容量．即　$v(f)\leqslant C(V_1,\bar{V}_1)$．

若对于任一个可行流 f^*，网络中有一个割集$(V_1^*,\bar{V}_1^*)$使 $v(f^*)=C(V_1^*,\bar{V}_1^*)$，则 f^* 必是最大流而$(V_1^*,\bar{V}_1^*)$必是 D 的所有割集中容量最小的一个，即最小割集．

定理 1　网络中的可行流 f 是最大流，当且仅当网络中不包含可增路．

证明：必要性

若网络中包含可增路，基于 P 的调整流 $v(\tilde{f})=v(f)+\delta(P)>v(f)$，则 f 不是最大流．

充分性：若网络中不包含可增路，往证 f 是最大流．取 $v_1=\{v_s\}$，若 $v_i\in V_1$，且 $f_{ij}<C_{ij}$，则将 V_1 扩充为 $V_1\cup\{v_j\}$，a_{ij}称非饱合弧；否则称非饱合弧．

若 $v_i\in V_1$，且 $f_{ji}>0$，将 V_1 扩充为 $V_1\cup\{v_j\}$，a_{ji}称为非零流弧；否则称为零流弧．

由于网络中无 f 可增路，$v_t\subset\bar{V}_1$，得割集$(V_1,\bar{V}_1)$，$f_{ij}=\begin{cases}C_{ij} & 若\ a_{ij}\in(V_1,\bar{V}_1)\\ 0 & 若\ a_{ij}\in(\bar{V},V_1)\end{cases}$　$\therefore v(f)=C(V_1,\bar{V}_1)$，从而 f 是最大流，证毕．

若 f 是最大流，该网络中必存在一割$(V_1,\bar{V}_1)$，使 $v(f)=C(V_1,\bar{V}_1)$，说明 f 是最大流，$(V_1,\bar{V}_1)$是最小割．

定理 2　任一网络中，最小割容量与最大流的流量相等．

证明略，该定理中作为检验最大流或最小割的手段．

二、Ford 和 FulKerson 标号法

1. 在介绍该算法之前，将 Ford 和 FulKerson 算法用于解决图 2-27.

为了记录可增路，从 v_s 开始，给每个顶点 v_i 规定一个标号 (v_x, δ_i)，v_x 表示从 v_s 到 v_i 可增路上 v_i 的前一个顶点，δ_i 是从 v_s 输送到 v_i 的超过现行流的额外流量，符合流量守恒和容量限制条件，令 v_s 标号为 $(0, +\infty)$.

对图 2-27 中，v_s 标号 $(0, +\infty)$，考察与顶点 v_s 关联的弧，对于弧 a_{s1}，$f_{s1}=C_{s1}=3$，a_{s1} 是饱合弧；不满足标号条件；在弧 a_{s2} 上，$f_{s2}=1$，$C_{s2}=3$，$f_{s2}<C_{s2}$. ∴ a_{s2} 为非饱合弧，给 a_{s2} 增加流量，它必不超过前一弧的额外流量，也不能超过该弧的过剩容量 $C_{s2}-f_{s2}=3-1=2$，即 $\delta_2=\min\{\delta_s,(c_{s2}-f_{s2})\}=\min\{+\infty, 3-1\}=2$，将 v_2 标号为 $(v_s, 2)$；检查与 v_2 关联的弧，在弧 a_{12} 上，$f_{12}=2>0$ 为非零流弧，是反向弧，故 $\delta_1=\min\{\delta_2, f_{12}\}=\min\{2,2\}=2$，记 v_1 标号为 $(v_2, 2)$；弧 a_{23} 上 $f_{23}=c_{23}=4$ 为饱合弧；在弧 a_{42} 上，$f_{42}=1>0$，为非零流弧，故 $\delta_4=\min\{\delta_2, f_{42}\}=\min\{2,1\}=1$，将 v_4 标为 $(v_2, 1)$；检查与 v_4 关联的弧，在弧 a_{34} 上，$f_{34}=1>0$，为非零流弧，故 $\delta_3=\min\{\delta_4, f_{34}\}=\min\{1,1\}=1$，将 v_3 标为 $(v_4, 1)$；在弧 a_{4t} 上，$f_{4t}=0<2=c_{4t}$，a_{4t} 为非饱合弧，故令 $\delta_t=\min\{\delta_4,(c_{4t}-f_{4t})\}=\min\{1,2\}=1$. 从而找出起点从 v_s 终点到 v_t 的一条可增路 (v_s, v_2, v_4, v_t)，其额外增加流量 $\delta=\delta_t=1$.

2. 可增路上流的调整.

f_{s2} 转变为 $f_{s2}+1=1+1=2$

f_{42} 转变为 $f_{42}-1=1-1=0$ （a_{42} 反向弧）

f_{4t} 转变为 $f_{4t}+1=0+1=1$ 其它弧流量不变

此新可行流的流量 $v(f)$，变为 $v(f)+\delta=4+1=5$，例如 2-28 可以重复一下前面的过程，寻找 f 可增路.

给 v_s 标号 $(0, +\infty)$，检查与 v_s 关联弧，在弧 $a_{s1}=1$ 上，$f_{s1}=c_{s1}=3$，是饱合弧；$f_{s2}=2<3=c_{s2}$，为非饱合弧，$\delta_2=\min\{\delta_s, c_{s2}-f_{s2}\}=\min\{+\infty, 1\}=1$，将 v_2 标为 $(v_s, 1)$；检查与 v_2 关联弧，在反向弧 a_{42} 中，f_{42} 为零流弧，顺向弧 a_{23} 中，$f_{23}=4=c_{23}$，是饱合弧，在反向弧 a_{12} 上，$f_{12}=2>0$，∴ $\delta_1=\min\{v_2, f_{12}\}=\min\{1,2\}=1$，将 v_1 标为 $(v_2, 1)$；检查与 v_1 关联的弧，在弧 a_{13} 上，$f_{13}=1=c_{31}$ 为饱合弧，标号过程停止，不存在从 v_s 到 v_t 的可增路，从而现行流 f 为最大流，停止搜索，最大流 $v(f)=5$.

标号集合 $V_1=\{v_s, v_2, v_1\}$，$v_s\in V_1$，且 $v_t\in \overline{V}_1$，$c(V_1, \overline{V}_1)=c_{13}+c_{23}=v(f)$，说明 f 是最大流，而 $(V_1, \overline{V}_1)$ 为最小割. $v(f)=f_{s1}+f_{s2}=f_{3t}+f_{4t}=5$. 开始提出的问题得到解决. 最小割集是制约总运输量关键，改善或提高最小割集运输能力将会使总的运输量提高.

Ford 和 Fulkerson 标号算法：

(1) 可行流 f_{ij} 对应值是 f_{ij}，若网络中没有给定 f，则 f 是零流；

(2) 发点 v_s 标号 $(0, +\infty)$，v_s 是已标号未检查的点，其它点是未标号点；

(3) 如果不存在已标号，而未检查的点，算法终止，最大流是 f；否则按标号先后次序取一个已标号而未检查点 v_i，对一切未标号点 v_j，若 (v_i, v_j) 为非饱合弧，$f_{ij}<c_{ij}$ 给 v_j 以标号 (v_i, δ_j)，$\delta_j=\min\{\delta_i, c_{ij}-f_{ij}\}$.

若 (v_j, v_i) 为非零流弧，$f_{ji}>0$，则给 v_j 以标号 (v_i, δ_j)，$\delta_j=\min\{\delta_i, f_{ji}\}$，$v_j$ 成为已标号而未检查点. 查完一切未标号点 v_j 后，v_i 成为已标号并检查过的点.

(4)如果收点 v_t 未标号，转(3)，否则转(5)

(5)从 v_t 开始，通过第一标号，逆向找可增路 P，并令 $f_{ij}=\begin{cases} f_{ij}+\delta_t & a_{ij}\text{顺向弧} \\ f_{ij}-\delta_t & a_{ij}\text{反向弧} \\ f_{ij} & \text{其它} \end{cases}$

去掉所有点的标号，转(2)

第五节　最小费用最大流模型

一、问题的提出

上一节讨论了寻求网络最大流的问题，在实际生活中，涉及流的问题时，人们考虑的不只是流量，还有"费用"的因素，本节介绍的最小费用最大流问题就是这类问题之一.

若上节图 2-27 中表示的是某油田石油运输管网最大输油量 5 桶/分钟，若给出输油网上各管道 a_{ij}输送石油的单位流量运费如图 2-29 试确定方案，使费用最小，输油量最大.

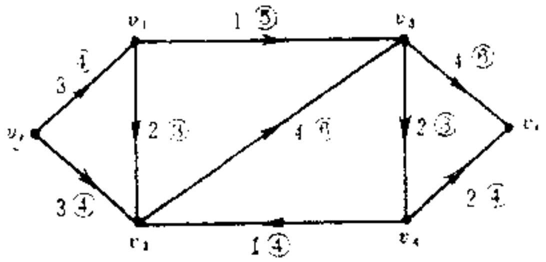


图 2-29　某油田石油运输管网(弧旁数字为：容量(费用))

一般在网络 D 上，对应每一弧 $a_{ij}\in A$，有一个容量 c_{ij}和一个单位流量的费用 $b_{ij}\geqslant 0$，求一个流量为 F 的可行流 f，使流的总输送费用

$b(f)=\sum\limits_{a_{ij}\in A} b_{ij}f_{ij}$为最小，其数学模型为

$$\min\sum_{a_{ij}\in A} b_{ij}f_{ij}\quad \sum f_{ij}-\sum f_{ji}=\begin{cases} F & i=s \\ 0 & \text{其它} \\ -F & i=t \end{cases}\quad F\leqslant \min C(V_1,\overline{V}_1),$$

$V_1\subset V$，等号若成立，F 即为最大流. 该模型即为最小费用最大流模型.

二、解法

1. 圈算法

定义 1：设网络 D 的一个可行流为 f，如果其流量 $v(f)=0$，则称其为循环流，此时费用为

$$b(f)=\sum_{a_{ij}\in A} b_{ij}f_{ij}$$

我们用类似求最大流的方法求最小费用最大流.

调整容量的流网络 $D'(f)$ 如下：D' 与 D 顶点集合相同；D 上弧 (v_i,v_j)，其上的流为 f_{ij}，容量为 c_{ij}，费用 b_{ij}，D' 中有两条弧 (v_i,v_j) 和 (v_j,v_i) 与之对应，弧 (v_i,v_j) 的容量为 $c_{ij}-f_{ij}\geqslant 0$，费用 b_{ij}；弧 (v_j,v_i) 的容量为 $f_{ij}\geqslant 0$，费用为 $-b_{ij}$，去掉所有容量为 0 的弧，如图 2-30.

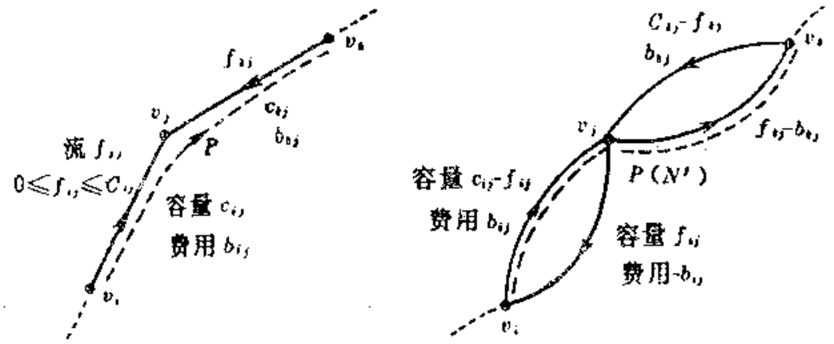


图 2-30　网络 D 中的弧与调整的网络 $D'(f)$ 中对应的弧

由定义可得，$D'(f)$ 中由 v_s 到 v_t 的一条向路 $P(D')$ 确定了 D 中一条 v_s 到 v_t 可增路 P，当沿着可增路 P 增加单位流量时，得到新的可行流 f'，此时增加的费用为

$$
\begin{aligned}
b(f') - b(f) &= \sum_{(v_i,v_j)\in P} b_{ij}(f'_{ij} - f_{ij}) - \sum_{(v_j,v_i)\in P} b_{ij}(f'_{ij} - f_{ij}) \\
&= \sum_{(v_i,v_j)\in P} b_{ij} - \sum_{(v_j,v_i)\in P} b_{ij}
\end{aligned}
$$

该式称为可增路 P 的"费用"，等于 $D'(f)$ 中自 v_s 到 v_t 费用权，或者说 $D'(f)$ 中自 v_s 到 v_t 的费用权为 x 的一条有向路确定了 N 中可增加流量 $\delta(P)$ 的一个新流 f'，其费用增加了 $x\cdot\delta(P)$；同样 $D'(f)$ 中费用权为 x 的有向圈 C，对应于 $D'(f)$ 中流量为 $\delta(c)$ 的一个循环流 $\bar{f}$，由此确定 D 中一个新的流 $f+\bar{f}$，其费用增加了 $x\cdot\delta(c)$.

定理 1：网络 D 中流量为 F 的流 f 是最小费用的，当且仅当 $D'(f)$ 中没有负费用的有向圈.

证明：必要性，假设 $D'(f)$ 中有负费用为 $-x$ 的有向圈 c，仿上面做法，在 D 中得到一个新的流 $f+\bar{f}$，其费用的增加为 $-x\cdot\delta(c)$，即减少了 $x\cdot\delta(c)$，与 f 是最小费用的矛盾.

充分性，反之，若 f 不是最小费用的，则存在一个费用比 f 小的流 f'，考虑 $f'-f$，由于 $v(f')=v(f)$，故 $f'-f$ 为循环流，对应 $D'(f)$ 中若干个有向圈 $c_1,c_2,\cdots c_q$，其上面流量分别为 $f_1,f_2,\cdots f_q$，循环流的 $f'-f$ 费用 $b(c_1)f_1+b(c_2)f_2+\cdots+b(c_q)f_q<0$，其中必有一有向圈 c_r，使 $b(c_r)f_r<0$，如图 2-31，即 $b(c_r)<0$，从而 c_r 为 $D'(f)$ 中负费用的有向圈矛盾.

圈算法具体步骤如下：

(1)利用最大流算法找出流量为 F 的流 f；

(2)构造 f 对应的调整容量的流网络 $D'(f)$；

(3)搜索 $D'(f)$ 中的负费用有向圈 C，若没有则停止；否则转(4)；

(4)在 c 中找出最大的循环流，并加到 D 上去，同时修改 $D'(f)$ 中 c 的容量，转(3).

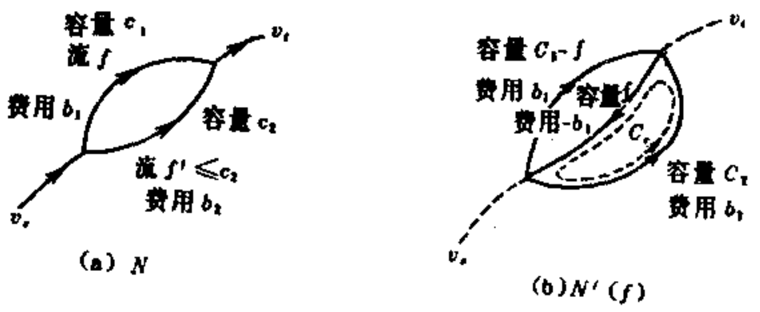


图 2-31

我们用"圈算法"解决图 2-29 所示石油运输网的最小费用流，由于最大流量 $v(f)=5$，该结果应是最小费用最大流．在图 2-28 的运输方案下，其费用

$$b(f)=\sum_{b_{ij}\in A} b_{ij}f_{ij}=4\times 3+5\times 1+6\times 4+4\times 2$$
$$+3\times 2+6\times 4+4\times 0+3\times 1+4\times 1=86$$

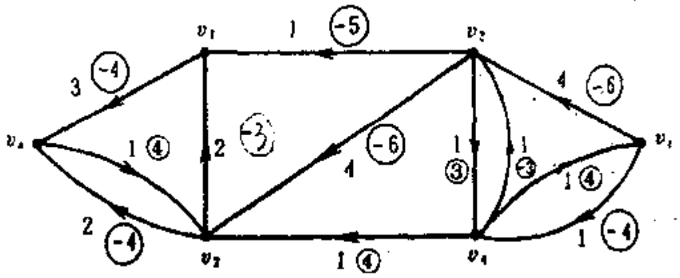


图 2-32　基于 f 调整容量的流网络(弧旁数字为：容量(运费))

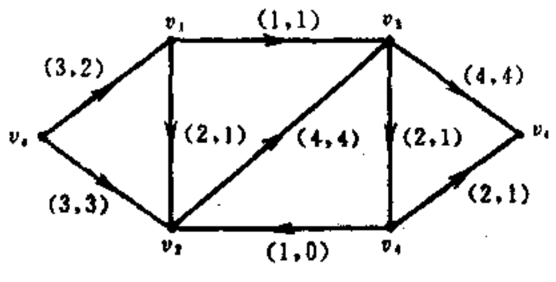


图 2-33　新流(弧旁数字为 (c_{ij},b_{ij}))

图 2-32 给出了 f 对应的调整容量的流网络 $D'(f)$，D' 中存在负费用有向圈 (v_s,v_2,v_1,v_s) 费用为 $4+(-3)+(-4)=-3$；容量为 $\min\{1,2,3\}=1$，将容量为 1 的该循环流加

到原来的流 f 上，得一新流，如图 2-33，总费用为 86+(−3)=83 新的调整网络如图 2-34，该网络无负费用有向圈，此流是最优的.

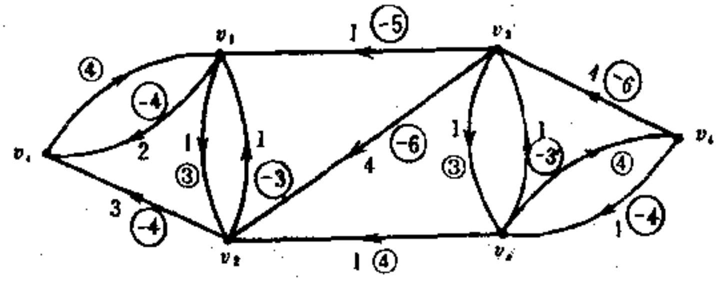


图 2-34　新流对应的调整容量的流网络

2. 迭加算法

我们在利用流 f 的可增路 p_1 求最大流的进行中，得到一新流 f'，流量增加为 $v(f)+\sigma(p_1)$. 若 f' 不是最小费用流，根据上面介绍的圈算法在其调整网络 $D'(f')$ 上一定找到负费用有向圈 c，将其循环流 $\bar{f}$ 加到原网络上，得到另外一流 $f'+\bar{f}$ 该流可以看成由 f 通过另外一条可增路 p_2 得到，而 p_2-p_1 对应于负费用有向圈 c，显然 p_2 的费用小于 p_1 的费用. 通过上述分析，试问：能否将求流量为 F 的最小费用流的圈算法化为在寻找可增路 p 的过程中定出我们需要的可增路 p_2，从而直接找出流量为 F 的最小费用流. 理论根据如下：

定理 2：设 f_1 是 D 中流量为 F 的最小费用流 f_2 是 $D'(f_1)$ 中沿最小费用的有向路的 p 单位流，那么 f_1+f_2 是 D 中流量为 $F+1$ 的最小费用流.

证明：假设 f_1+f_2 不是最优的，由定理 1 知在调整网络 $D'(f_1+f_2)$ 中有负费用的有向圈 c，因为流量为 F 的流 f_1 是最优的，所以 $D'(f_1)$ 中没有负费用的有向圈. 因此圈 c 是在流量从 F 增加到 $F+1$ 时产生的，故 c 中必有一弧 $(v_j,v_i)=e$，设其费用为 $-b$，使它对应于路 p 中的弧 (v_i,v_j) 如图 2-35.

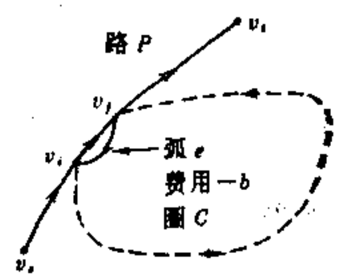


图 2-35　由负费用圈构造较低费用的路

若用 $c\setminus\{e\}$ 替换 p 上弧 (v_i,v_j)，则 p 费用的增加量为 $-b+(c\setminus\{e\})$ 的费用 $=c$ 的费用 <0，因此 p 不是 $D'(f_1)$ 中最小费用的有向路，矛盾. 说明 f_1+f_2 是流量为 $F+1$ 的最小费用流.

根据定理 2，我们可以从一个最小费用流（如零流）开始，沿 $D'(f)$ 中从 v_s 到 v_t 的最小费用的有向路，逐步增加流的值，直到流量为要求的 F 为止，若要求最小费用最大流，则一直到最大流求得为止.

迭加算法是 1962 年 Ford 和 Fulkerson 给出的，具体步骤如下：

(1)给定目标流量 F,或∞;从给定的最小费用流 f 开始;若未给,令 f 为零流.

(2)若 $v(f)=F$,停止运算,f 为最小费用流;否则转(3)

(3)在 $D'(f)$ 中寻找从 v_s 到 v_t 的最小费用有向路 p,沿 p 增加流 f 的流量直到 F,转(2);若不存在从 v_s 到 v_t 的最小费用的有向路 p,停止,f 就是最小费用最大流.

用迭加算法从零流开始,求图 2-29 最小费用最大流.如图 2-36

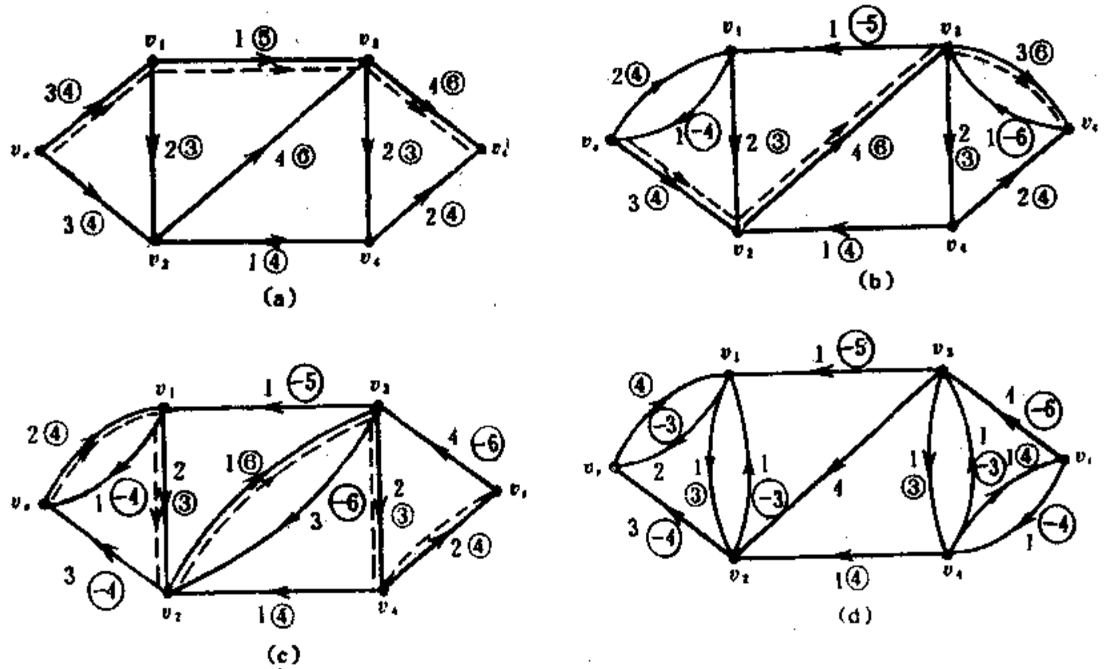


图 2-36 用迭加算法求最小费用最大流

(弧旁数字为容量(运费);虚线为短路)

首先构造调整容量的流网络 $D'(f)$,并求出从 v_s 到 v_t 的最短路(v_s,v_1,v_3,v_t),如图 2-36(a)$\delta=\min\{3,1,4\}=1$,将其添加到原网络 D 中去,进一步得新的调整网络,图 2-36(b)所示,按照迭加算法,依次得到流量为 1,4,5 的流,其对应的调整网络如图 2-36(b),(c),(d),最终得到最小费用为 83,流量为 5 的最小费用最大流.

第六节 匹配模型

问题的提出及解决

1. 最优分配问题

在分工问题中,工作人员可以做的各项工作效益未必一致,我们需要制定一个分工方案,使公司的总效益最大,这个问题称为最优分配问题.

构造一个图 G,G 有两个顶点集,$X=\{x_1,x_2,\cdots,x_n\}$,$Y=\{y_1,y_2,\cdots,y_n\}$分别表示工

作人员和工作，x_i，y_j 之间用边相连，其权 w_{ij}表示工作人员 x_i 做 y_j 工作时的效益，X 与 Y 内部没有边，所有边顶点分别在 X,Y 中，故图 G 是具有二分类(X,Y)的二部图(或偶图)，最优分配问题等价于在这个赋权图中寻找一个边集，使得任意两条边均不相邻，同时边集的总权最大，我们称为最大权匹配(或最优匹配)；二部图的赋权匹配问题又称为指派问题.

若 G 不是二部图，则问题转化为一般的赋权匹配问题. 在赋权匹配问题里，假定 G 是完全图，因为对于不在 G 里的边可以认为其权为 0；因此可以假定 G 有偶数个点，否则可以增加一个新点与其余所有的点都有边相连，这些边的权都为 0，类似地，对于二部图的赋权匹配问题可假定，X,Y 顶点的个数相等，且任意 x_i，y_j 之间有边相连，即为完全二部图.

设 M 是 G 的匹配，若匹配中的某条边与顶点 v 关联，则称 v 是饱合的，由于 $w_{ij}\geqslant 0$，故最优解总可以使任一顶点是 M 饱合的，从而其匹配是一个完美匹配.

最小权匹配问题，可以转化为权为 $c_{ij}=w-w_{ij}$的最大权匹配问题，其中 w 大于所有的 $w_{ij}(w_{ij}\geqslant 0)$

当所有的 $w_{ij}=1$ 时，问题即转化为不赋权匹配问题(基数匹配).

2. 基数匹配和覆盖

若图 G 中，匹配 M 是边最多的匹配，则称为 G 的最大匹配，完美匹配必是最大匹配，反之不然.

设 M 是 G 的匹配，G 的 M 交错路是指边在 $E\backslash M$ 和 M 中交错出现的路，M 可扩路是指其起点和终点都是 M 非饱合的 M 交错路.

定理 1：G 的匹配 M 是最大匹配当且权当 G 不包含 M 可扩路.

证明：设 M 是 G 的最大匹配，假设有一可扩路 $p_0=v_0,v_1,\cdots,v_{2m+1}$，令 $M'\subseteq E$ 为 $M'=(M\backslash\{(v_1,v_2),\{v_3,v_4\},\cdots,(v_{2m-1},v_{2m})\}\cup\{(v_0,v_1),(v_2,v_3),\cdots,(v_{2m},v_{2m+1})\}=(M\backslash P)\cup(P\backslash M)=M\triangle P$($M$ 与 P 的对称差)，则 M'是 G 的匹配，且$|M'|=|M+1|$；与 M 是最大匹配矛盾，故 G 中不包含 M 可扩路.

反之，当 G 不包含可扩路时，假设 M 不是最大匹配，则存在匹配 M'，$|M'|>|M|$. 令 $H=G[M\triangle M']$即 H 为由 M 与 M'的对称差，生成的 G 的子图. H 的任一顶点或是只与 M 或 M'中一条边关联，或是同时与 M 中一条边及 M'中一条边关联，于是 H 的每个分支或是一个边交错地属于 M 与 M'的长度为偶数的圈，或是一条交错地属于 M 与 M'的路，因为$|M'|>|M|$. H 中包含的 M'与边多于 M 的边，所以 H 中至少有一个分支 P，它所含的属于 M'的边比属于 M 的边多. 这个分支不可能是圈，只能是一条起点与终点均为 M 非饱合的 M 交错路，与 G 不包含 M 可扩路矛盾，证毕.

S 的邻集：对于图 G 的任一顶点集 S，其邻集为与 S 的顶点相邻的所有顶点的集，记为 $N(S)$.

设 G 是一个具有二分类(X,Y)的二部图，在分派问题中，能否存在这样一个分派方案使每个职员都有对应的工作可做？即在 G 中是否存在一个匹配，使它饱合 X 的每个顶点？定理 2 给出存在这种匹配的充要条件.

定理 2：设 G 是具有二分类(X,Y)的二部图，则 G 存在饱合 X 的每个顶点的匹配的充要条件是：对任何 $S\subseteq X$ 均有$|N(S)|\geqslant|S|$.

如果图 G 中所有顶点的度数都为 k，则称图 G 是 k 次正则图，由此上定理得如下推论.

推论：若 G 是 k 次正则二部图($k>0$)，则 G 有完美匹配.

"匈牙利算法"解决了二部图基数匹配问题，步骤如下：

(1)在 G 中任取一个匹配 M.

(2)若 M 饱和 X 的每个顶点，则算法终止，M 为完美匹配；否则设 u 是 X 中的 M 非饱合顶点，令 $S=\{u\}$，$T=\phi$.

(3)若 $N(s)=T$，由于 $|T|=|S|-1$，所以 $|N(s)|<|S|$，算法终止，由定理 2，不存在饱合 X 每个顶点的匹配，否则取 $y\in N(S)-T$.

(4)若 y 是 M 饱合时，设 $yz\in M$，用 $S\cup\{z\}$代替 S，$T\cup\{y\}$代替 T，转(3)(此时 $|T|=|S|-1$ 仍成立)；否则设 P 是 M 可扩路 $P(u,y)$，用 $M\triangle E(P)$代替 M，转(2)

例 1 已知 5 个职员 5 项工作的一个匹配如图 2-37(a)所示，求其最大匹配.

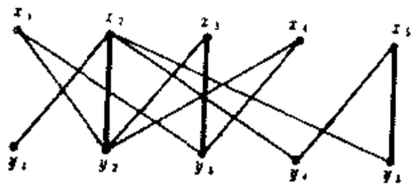


图 2-37(a)

解：

(1)取初始匹配 $M=\{x_2y_2,x_3y_3,x_5y_5\}$

(2)从非饱合点 x_1 开始生长外向交错树，找出一条 M 可扩路 $x_1y_2x_2y_1=P$ 如图 2-37(b)

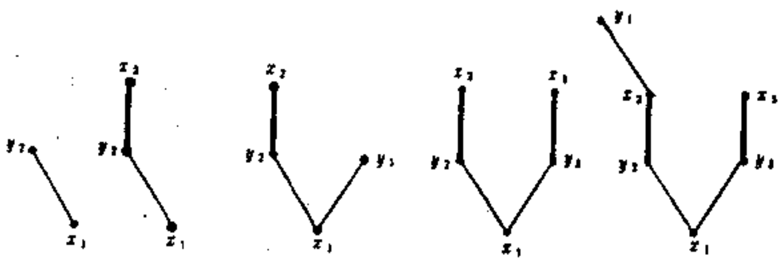


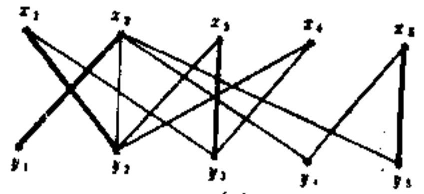
图 2-37(b)

(3)构造新的匹配 $M_1=M\triangle P=\{x_1y_2,x_2y_1,x_3y_3,x_5y_5\}$

(4)从非饱合点 x_4 开始生长外向交错树，未得到 M_1 可扩路，图 2-37(d)，故 M_1 是最大匹配，此时 $S=\{x_4,x_1,x_3\}$，$N(S)=\{y_2,y_3\}$，表面 G 不存在完美匹配.

图 G 的覆盖是指 V 的一个子集 k，使得 G 的每条边都至少有一个端点在 k 中. 如果 G

没有覆盖 k' 使 $|k'|<|k|$ 称 k 是最小覆盖.



(c)

图 2-37(c)

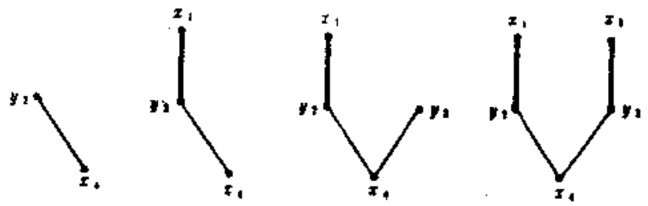


图 2-37(d)

例 2　图 2-38 表示一个居民小区,顶点表示交叉路口,边表示街道,现计划在某些路口安置消防设施,只有与路口直接相连的街道才能使用它们,为使所有的街道都有消防设施可用,在哪些路口安置设施最节省?

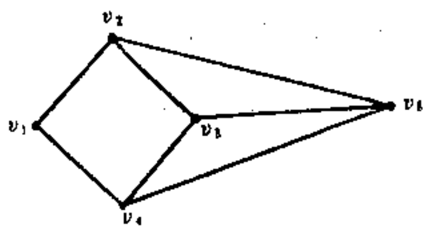


图 2-38

解:

(v_1,v_2,v_3,v_4),(v_1,v_3,v_4,v_5),(v_2,v_3,v_4),(v_1,v_3,v_5) 等都是最小覆盖,而 (v_2,v_3,v_4),(v_1,v_3,v_5) 等是最小覆盖,可以作为消防设施设置的地方.

定理 3:若 M^* 是二部图 G 的最大匹配,$\tilde{k}$ 是 G 的最小覆盖,则 $|M^*|=|\tilde{k}|$

3. 二部图的赋权匹配

对于有 $2n$ 个顶点的完全二部图,可以通过比较 $n!$ 个完美匹配的权去求最优匹配.当 n 很大时,该方法就无能为力了.

Kuhn—Munkres 算法解决了该问题.

定义 1:在 G 的顶点集 $V=X\cup Y$ 上定义一个实值函数 l,使得任何 $x\in X$ 和 $y\in Y$,均有 $l(x)+l(y)\geqslant w(x,y)$,称 $l(v)$ 为该二部图的一个**可行顶点标号**.

若 $E_l=\{(x,y)\mid(x,y)\in E,l(x)+l(y)=w(x,y)\}$,则称以 E_l 为边集的 G 的生成子图为 G 对应于可行顶点标号 l 的**相等子图**,记为 G_l.

$$\text{令}\quad l(x)=\begin{cases}\max\limits_{y\in Y} w(x,y) & x\in X\\ 0 & x\in Y\end{cases}$$

定理 4：设 l 是 G 的可行顶点标号，若 G_l 包含完美匹配 M^*，则 M^* 是 G 的最优匹配.

Kuhn－Munkres 算法

(1)从任一可行顶点标号 l 开始，确定 G_l，并在 G_l 中选取任一匹配 M；

(2)若 X 是 M 饱合的，则 M 是 G_l 的完美匹配，是 G 的最优匹配，算法终止；否则，在 G_l 中取一个 M 非饱合顶点 u，令 $S=\{u\}$，$T=\phi$.

(3)若 $N_{G_l}(S)\supset T$，转(4)；若 $N_{G_l}(S)=T$，计算

$$\alpha_l=\min_{\substack{x\in S\\ y\notin T}}\{l(x)+l(y)-w(x,y)\}\text{且由}$$

$$\hat{l}(v)=\begin{cases} l(v)-\alpha_l & v\in S \\ l(v)+\alpha_l & v\in T \\ l(v) & \text{其它} \end{cases}$$

给出可行顶点标号 $\hat{l}$，以 $\hat{l}$ 代替 l，以 $G_{\hat{l}}$ 代替 G_l.

(4)在 $N_{G_l}(S)\backslash T$ 中选择一个顶点 y，若 y 是 M 饱合的，并且 $yz\in M$，则用 $S\cup\{z\}$ 代替 S，用 $T\cup\{y\}$ 代替 T，转(3)；否则取 G_l 中一条 M 可扩路 $P(u,y)$，用 $M\triangle E(P)$ 代替 M，并转(2).

第七节　中国邮路模型

一、问题的提出

一个邮递员从邮局出发，到所管辖街道投递邮件，最后返回邮局，如果他必须走遍所辖的每条街道至少一次，那么应如何选择投递路线，使所走的路程最短，这个问题首先是我国学者管梅谷教授于 1962 年提出的，因此称为中国邮路问题.

中国邮路问题具有普遍意义，如公安执勤人员执行巡逻任务，企业的巡回检测等等都存在如何选择最短巡回路线的问题.

如果把投递区的街道用一条边 (v_i,v_j) 表示. 街道的长度用边权 $w(v_i,v_j)$ 表示，邮局、街道交叉口用点表示，那么一个投递区构成一个边权连通无向图，邮路问题用图论的语言来描述，就是在一个边权连通无向图中，怎样寻找一个回路 C，使得 C 经过每条边至少一次且 C 的长度最短. 因此邮路问题，既是一个与欧拉回路有关，而且也与最短路径有关的问题.

二、几种邮路问题及其解法

1. 最理想情况下邮路问题的解

连通赋权图上 G 上，一条过每边恰一次的链称为欧拉链，含闭欧拉链的图 G 为欧拉图，闭欧拉链称为欧拉回路，欧拉链或欧拉圈可以形象地描述为“一笔画”问题.

如果邮路 $G=(V,E)$ 是一个连通的欧拉图，则存在欧拉回路，这是最理想情况，每条边只需经过一次，回路的长度 $L(C)=\sum\limits_{(v_i,v_j)\in E}w(v_i,v_j)$ 是最小可能值，邮路的投递路线，也就是欧拉回路的遍历.

2. 次理想情况下邮路问题的解

如果邮路 G 具有欧拉通路，即只有两个结点 v_i 和 v_j 是奇次结点，这时从 v_i 出发可以经过每条边一次且仅一次而到达 v_j，从 v_j 回到 v_i 必须重复经过一些边，显然，要想使总的投递路线最短，必须使重复边的总长度最小，因此该问题成为寻找 v_i 和 v_j 之间最短路径问题，这种情况称为次理想情况. 其算法如下

(1)求出奇次结点 v_i 和 v_j 之间的最短路径 P.

(2)令 $G^*=G\cup P$

(3)G^* 为 E 图，穿行 G^* 的欧拉回路就是邮路投递路线的最优解.

例 1　图 2-39 表示一邮递路线，求出最优投递路线.

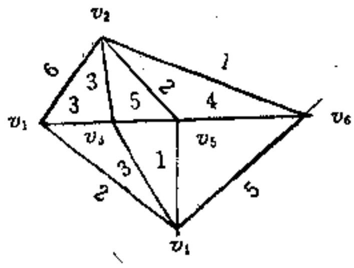


图 2-39

解：求出 v_1 到 v_6 的最短路径 $P=v_1,v_4,v_5,v_2,v_6$，其长度为 6，如图 2-40(a)

将 P 加到 G 上，得到图 G^*，如图 2-40(b)，最优投递路线即为 G^* 的欧拉回路.

$C=v_1v_2v_3v_5v_2v_6v_4v_5v_2v_6v_5v_4v_3v_1v_4v_1$

$$\text{总长度为 } L(C)=\sum_{(v_i,v_j)\in E^*} w(v_i,v_j)$$

$$=34+6=40$$

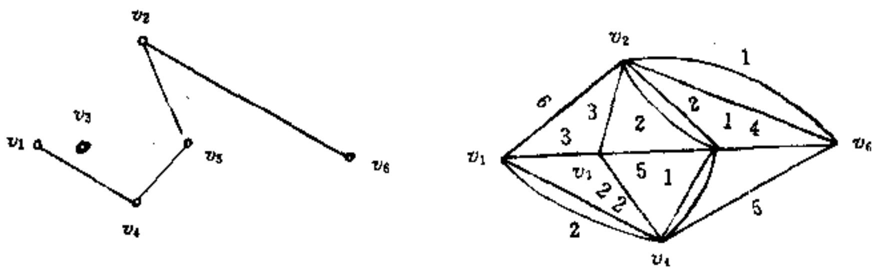


图 2-40

3. 一般情况下邮路问题的解

当邮递线路既非欧拉回路，又不具有欧拉通路时，即为一般情况，在这种情况下，必须重复更多的边，如何才能使投递路线的总长最小呢？下面的定理规定了两个最优路线的条件.

定理 1：设 G^* 是包含了连通边权无向图 G 的所有边的一个回路，则 G^* 具有最小长度的充分必要条件是：

(1)G 的每条边最多重复一次.

(2)在 G 的每个回路上，有重复边的长度和不超过回路长度的一半.

证明：必要性，首先设有些边重复 $n(n\geq 2)$ 次得到的图是欧拉图，如图 2-38(a)，(重复边用虚线表示)，将重复的边数去掉偶数条(最后必然只重复一次或零次)，重复边的两个端点的次数仍然是偶数，结果仍是欧拉图，如图 2-41(b)所示，所以最优的情况下边的重复不会多于一次.

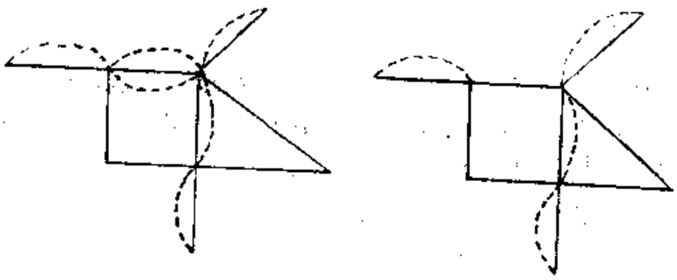


图 2-41

其次，我们考察 G 的任一回路，如图 2-42(a) 如果把重复的边都不重复，不重复的边都重复一次，则图(a)变成图(b)，这样每个结点的次数都改变了. 0 或 2，这种改变的结果并不影响结点的偶次性，图仍然保持欧拉图的性质，如果(a)重复边的长度和超过回路总长的一半，则(b)重复边的长度和必小于回路总长的一半，所以(b)是最优的.

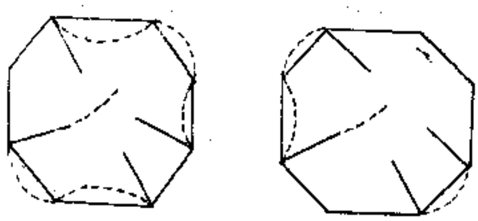


图 2-42

充分性：只要证明满足定理的两个条件的所有回路的长度均相等，即重复边的长度和相等即可.

设 G_1 和 G_2 是满足上述两个条件的回路，C 是 G_1 与 G_2 中复重边的环和，由边集 C 组成的网络的每个分支都是欧拉回路，因为对 G 的任一结点 v_1，G_1 和 G_2 中与 v 关联的重复边的数目的奇偶性相同，取决于 v 原来次数的奇偶性，如 v 原为奇次结点，则重复边的数目也为奇数，否则为偶数. 因此 C 是回路的边不相重复的并集，但在每一个回路上，G_1 和 G_2 的重复边长度都不超过回路长的一半，故只能相等且等于回路长的一半，因而 C 在 G_1 和 G_2 中重复边的长度相等，于是 G_1 和 G_2 中重复边的长度相等，即 G_1 与 G_2 的长度相等.

定理的证明给我们提供了一般情况下，寻求最优投递路线的方法，称为奇偶点作业法.

奇偶点作业法

1. 任给一个初始方案，使网络各结点的次数均为偶数，网络成为边权欧拉图.

2. 检查每一回路重复边的长度和是否不超过回路长度的一半，如是，则现行方案即为最优解，否则进行下一步.

3. 调整重复边，即将回路中重复的边改为不重复，不重复的边改为重复，返回第 2 步.

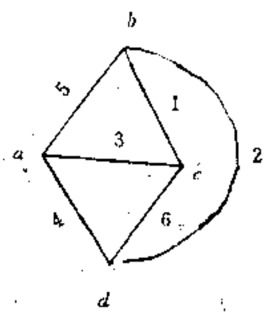


图 2-43

例 2　用奇偶点作业法，对图 2-43 所示的网络，求最优投递路线.

在 G 上增加两条边 (a,b)，(c,d).

检查回路 $abca$，重复边长度超过回路总长一半，调整重复边，欧拉图 G_1 变成欧拉图 G_2，如图(b)所示. 检查回路 $bcdb$，重复边长度超过回路长度一半，调整重复边，得到图 G_3，如图(c)所. 图 G_3 不存在重复边长度和超过回路长度一半的回路，因此 G_3 是最优解.

"奇偶点图上作业法"对回路较少的是可行的，当回路很多时，要检查所有的回路是难以做到的，Edmonds 在 1965 年提出了一种有效的方法，使中国邮路问题得到了较好的解决.

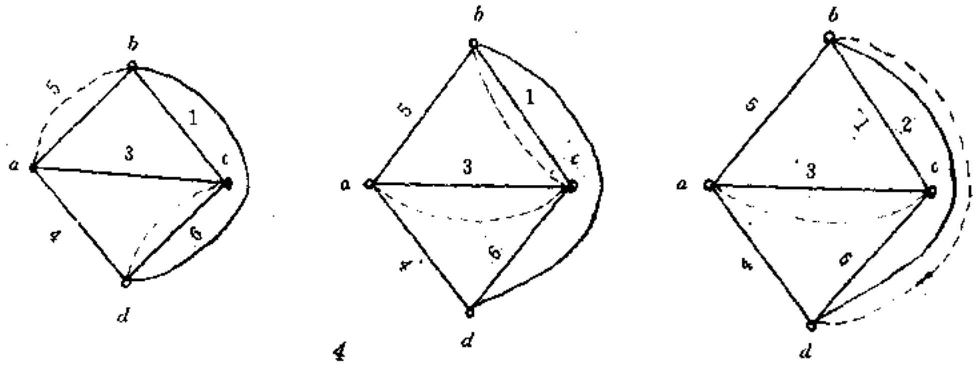


图 2-44

Edmonds 基本思想

(1)根据给定图 G，构造一个新图 G^*，G^* 中的顶点就是 G 中的所有奇点，并将 G^* 中任意两顶点都相连，此时 G^* 是一个完全图，G^* 中边 (v_i, v_j) 的权等于 G 中顶点 v_i 和 v_j 间的距离(即 v_i 到 v_j 的最短路长).

(2)在 G^* 中找一个最小权完全匹配 M(M 是 G^* 中具有如下性质的一个边集：G^* 中每个点恰与 M 中一条边关联，且 M 的权为最小).

(3)在 G 中将互相匹配的奇点用最短路径相连，由此得到 G 的最小新增边集.

例 3 用 Edmonds 算法求出图 2-45(a)中的邮递员问题.

根据奇点 v_1, v_2, v_3, v_5 构造一完全图 G^*，如图 2-45(b)所示，求 G^* 的最小权完美匹配 M，得 $M=\{(v_1, v_2), (v_3, v_5)\}$，其权为 $2+5=7$，在 G 中加入 (v_1, v_2) 和 (v_3, v_4, v_1, v_5) 两条最短路，使 G 变为 Euler 图，见图 2-43 其权值 $26+7=33$ 与"奇偶点图上作业法"的结果一样.

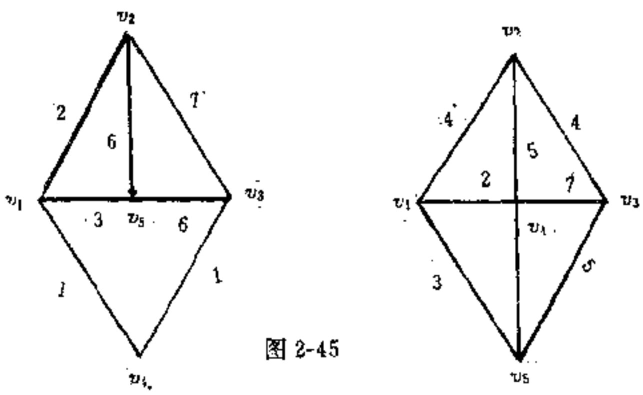


图 2-45

习题二

1. 用破圈法和避圈法找出图 2-46 中的一个生成树.

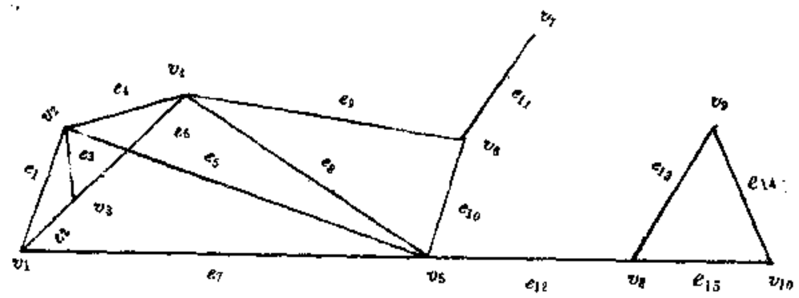
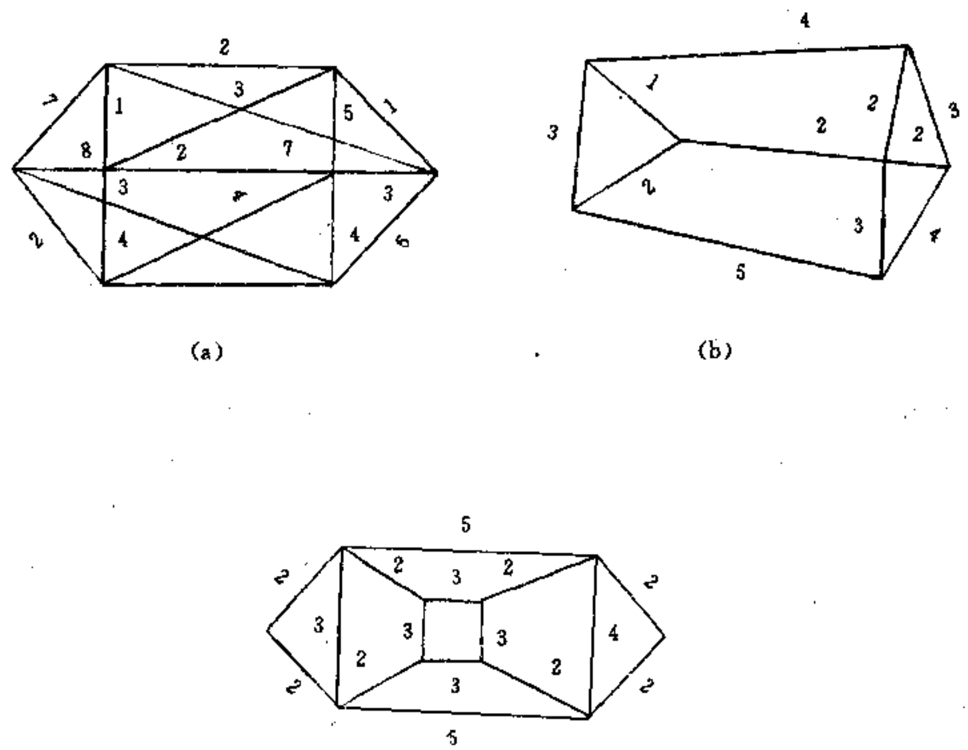


图 2-46

2. 用破圈法和避圈法求图 2-47 中各图的最小生成树.



(a) (b) (c)

图 2-47

3. 有九个城市 $v_1, v_2, \cdots, v_9$，公路网如图 2-48 所示，弧旁数字是该段公路的长度，有一批货物从 v_1 运到 v_9，问走哪条路最短？

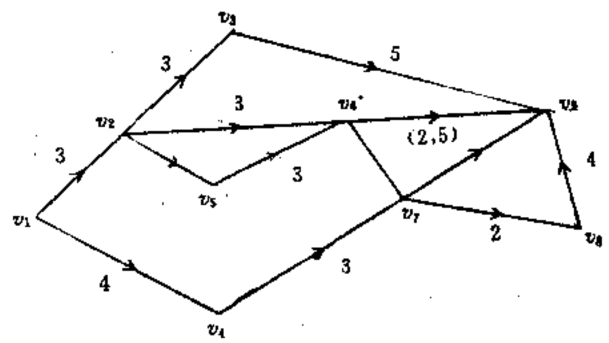


图 2-48

4. 用 Dijkstra 方法求图 2-49，从 v_1 到各点的最短路.

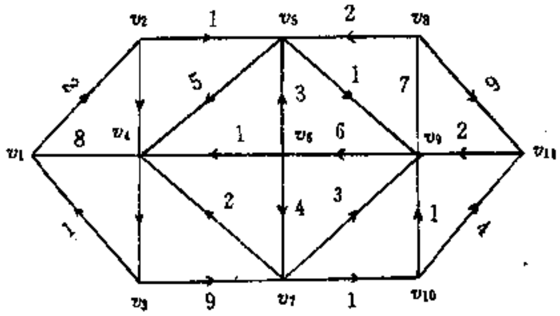


图 2-49

5. 求图 2-50 中所示的网络最大流(弧旁的数字是(c_{ij},f_{ij}))

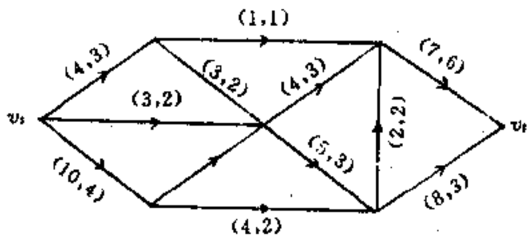


图 2-50

6. 求如图 2-51 所示的网络最大流.

7. 求如图 2-52 所示网络的最小费用最大流，每弧旁的数字是(b_{ij},c_{ij}).

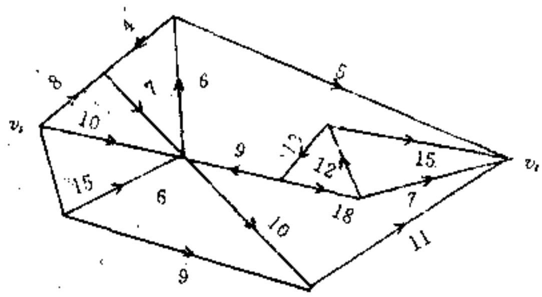


图 2-51

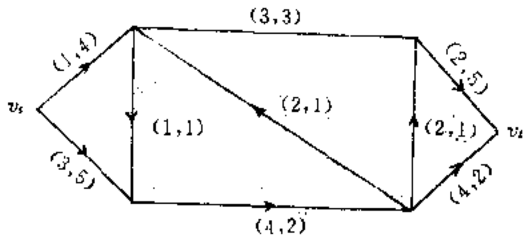


图 2-52

8. 求如图 2-53 所示中国邮递员问题.

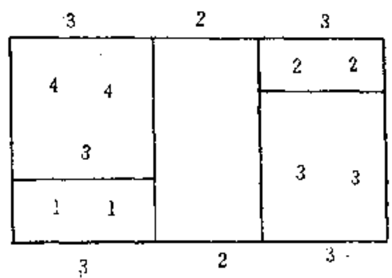


图 2-53

9. 设 G 是一个连通图,不含奇点,证明:从 G 中丢去任一条边后,得到的图仍是连通图.

10. 下述论断正确与否:可行流 f 的流量为零,即 $v(f)=0$,当且仅当 f 是零流.

11. 假设两个国家 A 与 B 处于战争状态,A 可有选择地对 B 的领土进行轰炸,B 的领土中包含一个桥与路的网络,A 国想破坏 B 国中的物流通,试用流的概念去解释对于 A 国来说,炸桥比炸路更有益处.

第三章　数值分析建模

在生产和科研实践所碰到的大量的函数，相当部分是通过实验和观测得到的. 虽然它们的函数关系是客观存在的，但是却不知道其具体的解析表达式，而只能得到离散的数据点：(x_i, y_i)，$i=0,1,\cdots,n$. 因此，希望能对这组数据作分析、处理从而得到比较简单的表达式近似地给出整体上的描述. 又由于测定过程中不可避免会产生误差，加之处理数据时所采用方法的近似性，故在求解时又必须考虑这些误差对结果的影响，作好理论分析（收敛性、稳定性、误差分析等）. 据此合理建立实际系统数学模型的方法称为数值逼近方法.

本章主要介绍插值法建模及曲线拟合法建模.

第一节　拉格朗日插值及模型

通过实验测得函数 $y=f(x)$ 的两两互异的点 $x_0, x_1, \cdots, x_n$ 的函数值，即得出如下的数据表

表 3-1

x	x_0	x_1	……	x_i	……	x_n
y	y_0	y_1	……	y_i	……	y_n

求自变量 x 与因变量 y 的一个近似的函数关系式 $y=\varphi(x)$，使其满足：

$$y_i = \varphi(x_i) \quad (i = 0, 1, \cdots, n) \tag{3.1}$$

其中　$x_0, x_1, \cdots, x_n$ 称为插值结点.

$\varphi(x)$ 称为插值函数.

条件式(3.1)称为插值条件. $[\min\{x_i\}, \max\{x_i\}]$ 称为插值区间，记作 $[a,b]$.

插值法是寻求近似函数的方法之一. 如果 x 不是结点，可计算 $\varphi(x)$ 在 x 点的值作为原来函数 $f(x)$ 在此点的近似值.

从几何上看，插值法就是要求一条曲线 $y=\varphi(x)$，它通过已知的 $n+1$ 个点 (x_i, y_i) $(i=0,1,\cdots,n)$，并用 $\varphi(x)$ 近似表示 $f(x)$（图 3-1）.

在用插值法寻求近似函数的过程中，对函数的类型可有不同的选取. 考虑到代数多项式结构简单，并有良好的分析性质，便于数值计算和理论分析，所以常用代数多项式作为插值函数. 本节主要介绍多项式插值，样条插值.

一、拉格朗日插值

1. 问题的提出

设 $y=f(x)$ 是表 3-1 所表示的函数，求次数不超过 n 的多项式 $p_n(x)$，使其满足插值

条件：

$$p_n(x_i) = y_i (i = 0,1,\cdots,n) \qquad (3.2)$$

这就是拉格朗日插值问题. 其中 $p_n(x)$称为 $f(x)$的插值多项式.

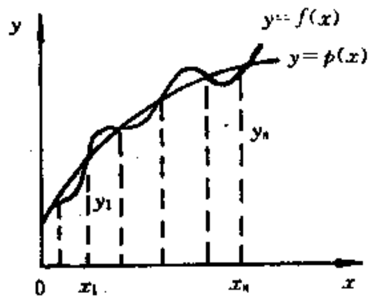


图 3-1

2. 插值多项式的存在性和唯一性

定理 1　满足插值条件(3.2)的次数不超过 n 的多项式存在且唯一.

证:设所求的插值多项式为

$$p_n(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n \qquad (3.3)$$

则由插值条件(3.2)得关于 $a_0, a_1, \cdots, a_n$ 的线性方程组

$$\begin{cases} a_0 + a_1x_0 + a_2x_0^2 + \cdots + a_nx_0^n = y_0 \\ a_0 + a_1x_1 + a_2x_1^2 + \cdots + a_nx_1^n = y_1 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_0 + a_1x_n + a_2x_n^2 + \cdots + a_nx_n^n = y_n \end{cases} \qquad (3.4)$$

它的系数行列式是范德蒙行列式

$$\begin{vmatrix} 1 & x_0 & x_0^2 & \cdots & x_0^n \\ 1 & x_1 & x_1^2 & \cdots & x_1^n \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & x_n & x_n^2 & \cdots & x_n^n \end{vmatrix} = \prod_{j>i}(x_j - x_i) \neq 0$$

故由克莱姆法则,方程组(3.4)的解存在且唯一. 定理 1 的结论成立. 证毕.

3. 拉格朗日插值多项式

为了构造满足条件(3.2)的插值多项式,首先求满足条件:

$$L_i(x_j) = \begin{cases} 0, & j \neq i \\ 1, & j = i \end{cases} \qquad (j = 0,1,\cdots,n) \qquad (3.5)$$

的 n 次多项式 $L_i(x)$　$(i=0,1,\cdots,n)$.

由式(3.5)可知,$L_i(x)$有 n 个零点 $x_0,\cdots,x_{i-1},x_{i+1},\cdots,x_n$,故存在常数 c,使

$$L_i(x) = c(x - x_0)\cdots(x - x_{i-1})\cdots(x - x_{i+1})\cdots(x - x_n),$$

此外由 $L_i(x_i)=1$,得

$$c = \frac{1}{(x_i - x_0)\cdots(x_i - x_{i-1})(x_i - x_{i+1})\cdots(x_i - x_n)}$$

即

$$L_i(x) = \frac{(x - x_0)\cdots(x - x_{i-1})(x - x_{i+1})\cdots(x - x_n)}{(x_i - x_0)\cdots(x_i - x_{i-1})(x_i - x_{i+1})\cdots(x_i - x_n)}$$

$$= \prod_{\substack{j=0 \\ j\neq i}}^{n} \frac{(x - x_j)}{(x_i - x_j)} \qquad (i = 0,1,\cdots,n) \qquad (3.6)$$

称(3.6)式为**拉格朗日插值基函数**.

由 $L_i(x)$可得

$$p_n(x)=y_0l_0(x)+y_1L_1(x)+\cdots+y_nL_n(x)=\sum_{i=0}^{n}y_iL_i(x) \tag{3.7}$$

就是满足插值条件(3.2)的拉格朗日插值问题的解，称式(3.7)为拉格朗日插值多项式.

特殊地，由两个数据点(x_0,y_0),(x_1,y_1)构造的拉格朗日插值多项式为

$$p_1(x)=y_0\frac{x-x_1}{x_0-x_1}+y_1\frac{x-x_0}{x_1-x_0} \tag{3.8}$$

从几何上看，$y=p_1(x)$为过两点(x_0,y_0),(x_1,y_1)的直线. 故(3.8)式又叫线性插值. 由三个数据点(x_0,y_0),(x_1,y_1),(x_2,y_2)构造的插值多项式为

$$p_2(x)=y_0\frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)}+y_1\frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)}+y_2\frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)} \tag{3.9}$$

又称(3.9)式为抛物插值.

4. 插值余项

把$R_n(x)\stackrel{\triangle}{=}f(x)-p_n(x)$称为插值余项. 关于余项有以下的定理.

定理 2 设$f(x)$在$[a,b]$上存在$n+1$阶导数，且$f(x_0)=y_0,f(x_1)=y_1,\cdots,f(x_n)=y_n$;则对区间$[a,b]$上的任何$x$，都存在$\xi$($\xi$依赖$x$)$\in(a,b)$，使得

$$R_n(x)=f(x)-p_n(x)=\frac{f^{n+1}(\zeta)}{(n+1)!}w_{n+1}(x), \tag{3.10}$$

其中
$$w_{n+1}(x)=(x-x_0)(x-x_1)\cdots(x-x_n)$$

证：当$x=x_i$时，3-10 显然成立.

当$x\neq x_i$时，由于$R_n(x)$满足$R_n(x_i)=f(x_i)-p_n(x_i)=0\quad(i=0,1,\cdots,n)$

故设
$$R_n(x)=k(x)(x-x_0)\cdots(x-x_n) \tag{3.11}$$

为确定$k(x)$，作辅助函数

$$\varphi(t)=f(t)-p_n(t)-k(x)(t-x_0)\cdots(t-x_n)$$

$\varphi(t)$有$n+1$阶导数，且在$[a,b]$上$\varphi(t)$有$n+2$个零点$x_0,x_1,\cdots,x_n,x$. 由罗尔定理依次类推知：$\varphi^{(n+1)}(t)$在(a,b)内至少有一点ξ，使

$$\varphi^{(n+1)}(\xi)=f^{(n+1)}(\xi)-(n+1)!k(x)=0$$

于是
$$k(x)=\frac{f^{(n+1)}(\xi)}{(n+1)!}$$

把上式代入(3.11)即可得出公式(3.10). 证毕

利用公式(3.10)可给出多项式$p_n(x)$逼近$f(x)$的误差估计，同时拉格朗日余项定理在理论上有很重要的价值，如余项公式只有在$f(x)$的$n+1$阶导数存在时才能应用，因此$f(x)$的光滑性较差时，多项式插值的效果就差. 再如，由于余项中含有因子$(x-x_0)\cdots(x-x_n)$，当点$x\in[\min\limits_{0\leqslant i\leqslant n}\{x_i\},\max\limits_{0\leqslant i\leqslant n}\{x_i\}]$时，插值效果好些，插值过程叫内插，否则称为外推. 一般说来，内插优于外推.

例 1 已知$\sqrt{100}=10$,$\sqrt{121}=11$,$\sqrt{144}=12$，分别用线性插值和抛物插值求$\sqrt{125}$，

并估计截断误差.

解：

①线性插值：选 $x_0=121, x_1=144$ 为插值结点

$$p_1(x)=11\frac{x-144}{121-144}+12\frac{x-121}{144-121}$$

$$\sqrt{125}\approx p_1(125)=11.173\,91$$

②抛物插值：

$$p_2(x)=10\frac{(x-121)(x-144)}{(100-121)(100-144)}+11\frac{(x-100)(x-144)}{(121-100)(121-144)}+12\frac{(x-100)(x-121)}{(144-100)(144-121)}$$

$$\sqrt{125}\approx p_2(125)=11.181\,07$$

③误差估计：$f(x)=\sqrt{x}$，$f'(x)=\frac{1}{2\sqrt{x}}$，$f''(x)=-\frac{1}{4x\sqrt{x}}$，即 $p_1(125)$ 的截断误差：

$$|R_1(125)|=\left|\frac{f''(\xi)}{2!}(125-121)(125-144)\right|=\frac{1}{8\xi\sqrt{\xi}}4\cdot 9\leqslant\frac{4\cdot 9}{8\cdot 121\cdot 11}$$
$$=0.007\,14\qquad(121<\xi<144)$$

$$|R_2(125)|=\left|\frac{f'''(\xi)}{3!}(125-100)(125-121)(125-144)\right|$$
$$=\left|\frac{3}{3!\cdot 8\xi^2\sqrt{\xi}}(125-100)(125-121)(125-144)\right|$$
$$\leqslant 0.001\,19\qquad(100<\xi<144)$$

二、丙烷导热系数的估计

丙烷的导热系数是化工生产中值得注意的量，而且常常需要在不同温度及压力下的导热系数，然而我们不可能也没有必要过细地实验测量，经过实验，给出数据如下：

表 3-2

T(℃)	P(kN/m²)	K(W/mK)	T(℃)	P(kN/m²)	K(W/mK)
68	$9.798\,1\times10^3$	0.084 8	106	$9.798\,1\times10^3$	0.069 6
	13.324×10^3	0.089 7		14.277×10^3	0.075 3
87	$9.007\,8\times10^3$	0.076 2	140	$9.656\,3\times10^3$	0.061 1
	13.355×10^3	0.080 7		12.463×10^3	0.065 1

表中的 T、P、K 分别表示温度、压力、导热系数，并假设在这个范围内导热系数近似地随压力线性变化，求压力是 $10.13\times10^3 kN/m^3$ 和 99℃温度下的导热系数.

解：从表面上看，这是一个二元函 $K(T,P)$ 的插值问题，即要求同时对压力和温度进行内插，但考虑到该问题假设中导热系数与压力的线性关系，将这个插值问题分两部分进行.

首先，利用表 3-2 中 P 和 k 的数据，分别求出 68℃，87℃，106℃，140℃时压力都是 $10.13\times10^3 kN/m^2$ 下的一组导热系数.

由线性插值求得 68℃时，压力在 $10.13\times10^3 kN/m^2$ 下的导热系数为：

$$K = 0.0848\frac{10.13-13.324}{9.7981-13.324} + 0.0897\frac{10.13-9.7981}{13.324-9.7981}$$
$$= 0.0853 W/mK$$

类似地求出在 87℃，106℃，140℃的导热系数，得出如下的表：

表 3-3

T(℃)	68	87	106	140
K(W/mK)	0.085 3	0.077 4	0.069 9	0.061 8

其次由数表 3-3 对温度进行内插，由拉格朗日插值多项式得温度 $T=99$℃的导热系数

$$K = 0.0853\frac{(99-87)(99-106)(99-140)}{(68-87)(68-106)(68-140)}$$
$$+ 0.0774\frac{(99-68)(99-106)(99-140)}{(87-68)(87-106)(87-140)}$$
$$+ 0.0699\frac{(99-68)(99-87)(99-140)}{(106-68)(106-87)(106-140)}$$
$$+ 0.0618\frac{(99-68)(99-87)(99-106)}{(140-68)(140-87)(140-106)}$$
$$= 0.0725 W/mK$$

下面是用 Mathematica 软件包，对丙烷导热系数问题数值实验的过程，从中可以得知所有插值过程的插值函数. 利用这些插值函数，能够对插值区间内的不同压力值和温度下的导热系数进行估计.

```
In[1]:=a={{9 798.1,0.084 8},{13 324,0.089 7}};b={{9 007.8,0.076 2},{13
      355,0.080 7}};
In[2]:=c={{9 791.8,0.069 6},{14 277,0.075 3}};d={{9 656.3,0.061 1},{12
      463,0.065 1}};
In[3]:=x=.;p1[x]=InterpolatingPolynomial[a,x]
out[3]=0.071 183 4+0.000 001 389 72x
In[4]:=p1[10 130]
out[4]=0.085 261 2
In[5]:=x=.;p2[x]=InterpolatingPolynomia[b,x]
out[5]=0.066 875 6+0.00 000 103 515x
In[6]:=p2[10 130]
```

```
out[6]=0.077 361 6
In[7]:=x=.;p3[x]=InterpolatingPolynomial[c,x]
out[7]=0.057 156 1+0.000 001 270 85x
In[8]:=p3[10 130]
out[8]=0.069 9
In[9]:=x=.;p4[x]=InterpolatingPolynomial[d,x]
out[9]=0.047 338 1+0.000 001 425 16x
In[10]:=p4[101 30]
out[10]=0.061 775 1
In[11]:=e={{68,0.085 3},{87,0.077 4},{106,0.069 9},{140,0.061 8}};
In[12]:=x=.;f[x]=InterpoLatingPoLynomiol[e,x]
out[12]=0.095 958 1+0.000 242 845x-0.000 008 141 79x^2+3.331 731 0^-8x^3
In[13]:=f[99]
out[13]=0.072 529 9
```

第二节　牛顿插值公式及模型

一、牛顿插值公式

利用拉格朗日插值公式时，如果需要增加一个插值节点并求出新的插值函数时，所有的基函数都要重新计算，以前的计算结果白白浪费．牛顿插值公式，可以利用以前计算的结果且临时增加新的结点，它要用到差商的概念．

1. 差商及其基本性质

①定义：设$(x_i,f(x_i))$，$i=0,1,\cdots,n$ 是 $f(x)$的 $n+1$ 对数据点，把

$f[x_0,x_1]=\dfrac{f(x_1)-f(x_0)}{x_1-x_0}$称为函数 $f(x)$在 x_0，x_1 点的一阶差商．

$f[x_0,x_1,x_2]=\dfrac{f[x_0,x_2]-f[x_0,x_1]}{x_2-x_1}$称为函数 $f(x)$在 x_0,x_1,x_2 点的二阶差商．

一般地 $n-1$ 阶差商的差商．

$$f[x_0,x_1,\cdots,x_n]=\frac{f[x_0,x_1,\cdots,x_{n-2},x_n]-f[x_0,\cdots,x_{n-2},x_{n-1}]}{[x_n-x_{n-1}]}$$

称为 $f(x)$在 $x_0,\cdots,x_n$ 点的 n 阶差商．

②差商的性质

1°　差商可以表示为函数值的线性组合，即

$$f[x_0,x_1,\cdots,x_n]=\sum_{i=0}^{n}a_if(x_i),$$

其中

$$a_i=\frac{1}{(x_i-x_0)\cdots(x_i-x_{2-1})(x_i-x_{i+1})\cdots(x_i-x_n)},$$

性质 1°也可叙述为：差商与结点的排列次序无关，即

$$f[x_0,x_1,\cdots,x_n]=f[x_1,x_2,\cdots,x_{n-1},x_0,x_n]=\cdots=f[x_1,x_2,\cdots,x_n,x_0].$$

2° $f[x_0,x_1,\cdots,x_n]=\dfrac{f[x_1,\cdots,x_n]-f[x_0,\cdots,x_{n-1}]}{x_n-x_0}$

证：因 $f[x_1,x_2,\cdots,x_{n-1},x_0,x_n]=\dfrac{f[x_1,\cdots,x_n]-f[x_1,\cdots,x_{n-1},x_0]}{x_n-x_0}$，再由性质 1°即证性质 2°.

3° 若 $f(x)$在(a,b)内存在 n 阶导数，结点 $x_i\in[a,b]$$(i=0,1,\cdots,n)$，则至少存在一点 $\xi\in[a,b]$，使

$$f[x_0,x_1,\cdots,x_n]=\frac{f^{(n)}(\xi)}{n!}$$

由该性质知，n 次多项式的 n 阶差商为常数. 因此，如果一个函数的 n 阶差商接近于常数，那么用 n 次多项式去近似它是合适的.

③差商表

利用性质 2°的公式，求 $f(x)$的各阶差商，得如下差商表.

表 3-4

x_0—$f(x_0)$				
	$f[x_0,x_1]$			
x_1—$f(x_1)$		$f[x_0,x_1,x_2]$		
	$f[x_1,x_2]$		$f[x_0,x_1,x_2,x_3]$	
x_2—$f(x_2)$		$f[x_1,x_2,x_3]$		$f[x_0,x_1,x_2,x_3,x_4]$
	$f[x_2,x_3]$		$f[x_1,x_2,x_3,x_4]$	
x_3—$f(x_3)$		$f[x_2,x_3,x_4]$		
	$f[x_3,x_4]$			
x_4—$f(x_4)$				

2. 牛顿插值多项式

由 $f[x_0,x]=\dfrac{f(x)-f(x_0)}{x-x_0}$ 得

$$f(x)=f(x_2)+f[x_0,x](x-x_0) \tag{3.12}$$

同理由 $f[x_0,x_1,x]=\dfrac{f[x_0,x]-f[x_0,x_1]}{x-x_1}$ 得

$$f[x_0,x]=f[x_0,x_1]+f[x_0,x_1,x](x-x_1) \tag{3.13}$$

……………

如此继续，由 $f[x_0,x_1,\cdots,x_n,x]=\dfrac{f[x_0,x_1,\cdots,x_{n-1},x]-f[x_0,x_1,\cdots,x_n]}{x-x_n}$得

$$f[x_0,x_1,\cdots,x_{n-1},x]=f[x_0,x_1,\cdots,x_n]+f[x_0,x_1,\cdots,x_n,x](x-x_n) \tag{3.14}$$

依次把式(3.13)…(3.14)代入(3.12)最后得

$$\begin{aligned}f(x)=&f(x_0)+f[x_0,x_1](x-x_0)+f[x_0,x_1,x_2](x-x_0)(x-x_1)+\cdots\\&+f[x_0,x_1,\cdots,x_n](x-x_0)\cdots(x-x_{n-1})\\&+f[x_0,x_1,\cdots,x_n,x](x-x_0)\cdots(x-x_n)\end{aligned} \tag{3.15}$$

记 $P_n(x)=f(x_0)+f[x_0,x_1](x-x_0)+\cdots+f[x_0,x_1,\cdots,x_n](x-x_0)\cdots(x-x_{n-1})$ (3.16)

则 $$f(x)=P_n(x)+f[x_0,x_1,\cdots,x_n](x-x_0)\cdots(x-x_n)$$

显然 $P_n(x)$满足插值条件(3.2)，且是次数不超过 n 的多项式，因此仍是拉格朗日插值问题的解，称为牛顿插值多项式. $P_n(x)$的系数是差商表 3-4 中用横线标出的各阶差商.

6B

由插值多项式的唯一性，它与拉格朗日插值多项式是等价的. 因此牛顿插值多项式的余项为(余项式对导数不存在的情况仍适用)

$$R_n(x) = f[x_0,x_1,\cdots,x_n](x-x_0)\cdots(x-x_n) \tag{3.17}$$

由余项式(3.10)和(3.17)等价，可证明差商的性质 3°.

实际应用中，需要增加一个结点时，式(3.16)常被写成如下递推公式

$$P_n(x) = P_{n-1}(x) + f[x_0,x_1,\cdots,x_n](x-x_0)\cdots(x-x_{n-1})$$

把结点按由大到小的顺序排列时，

$$P_n(x) = f(x_n) + f[x_n,x_{n-1}](x-x_n) + \cdots + f[x_n,x_{n-1},\cdots,x_1,x_0](x-x_n)\cdots(x-x_1) \tag{3.18}$$

例 1　给出 $f(x)$ 的函数表(见表 3-5)，求 4 次牛顿插值多项式，并由此计算 $f(0.596)$ 的近似值.

解：首先根据给定的函数表造出差商表.

表 3-5

x	$f(x)$	一阶差商	二阶差商	三阶差商	四阶差商	五阶差商
0.40	0.410 75					
0.55	0.578 15	1.116 00				
0.65	0.696 75	1.186 00	0.280 00			
0.80	0.888 11	1.275 73	0.358 93	0.197 33		
0.90	1.026 52	1.384 10	0.433 48	0.213 00	0.031 34	
1.05	1.253 82	1.515 33	0.524 93	0.228 63	0.031 26	−0.000 12

从差商表看到四阶差商近似为常数，故取 4 次多项式 $P_4(x)$ 作为 $f(x)$ 的近似即可.

$$\begin{aligned} P_4(x) &= 0.410\,75 + 1.116(x-0.4) + 0.28(x-0.4)(x-0.55) \\ &\quad + 0.197\,33(x-0.4)(x-0.55)(x-0.65) \\ &\quad + 0.031\,34(x-0.4)(x-0.5)(x-0.65)(x-0.8) \end{aligned}$$

于是

$$f(0.596) \approx P_4(0.596) = 0.631\,95$$

截断误差

$$|R_4(x)| \approx |f[x_0,\cdots,x_5]w_5(0.596)| \leqslant 3.63\times10^{-9}$$

二、等距结点的牛顿插值公式

设 $(x_i,f_i)(i=0,1,\cdots,n)$ 的结点 $x_i=x_0+ih$，这里 h 叫步长，是常数，$f_i=f(x_i)$.

1. 差分的定义及性质

定义：$\triangle f_i \xlongequal{\triangle} f_{i+1}-f_i$ 和 $\nabla f_i \xlongequal{\triangle} f_i-f_{i-1}$ 分别称为 $f(x)$ 在点 x_i 处的一阶向前差分和一阶向后差分.

$\triangle^2 f_i \xlongequal{\triangle} \triangle f_{i+1}-f_i$ 和 $\nabla^2 f_i \xlongequal{\triangle} \nabla f_i-\nabla f_{i-1}$ 分别称为 $f(x)$ 在点 x_i 处的二阶向前差分和二阶向后差分.

把$\triangle^m f_i \xlongequal{\triangle} \triangle^{m-1} f_{i+1} - \triangle^{m-1} f_i$和$\nabla^m f_i \xlongequal{\triangle} \nabla^{m-1} f_i - \nabla^{m-1} f_{i-1}$分别称为$f(x)$在$x_i$处的$m$阶向前差分和$m$阶向后差分.

表 3-6　差分表

f_i	一阶差分	二阶差分	三阶差分	四阶差分	…
f_0	$\triangle f_0, \nabla f_1$	$\triangle^2 f_0, \nabla^2 f_2$	$\triangle^3 f_0, \nabla^3 f_3$	$\triangle^4 f_0, \nabla^4 f_4$	…
f_1	$\triangle f_1, \nabla f_2$	$\triangle^2 f_1, \nabla^2 f_3$	$\triangle^3 f_1, \nabla^3 f_4$	…	
f_2	$\triangle f_2, \nabla f_3$	$\triangle^2 f_2, \nabla^2 f_4$	…	$\triangle^4 f_{n-4}, \nabla^4 f_n$	
f_3	$\triangle f_3, \nabla f_4$	…	$\triangle^3 f_{n-3}, \nabla^3 f_n$		
f_4	…	$\triangle^2 f_{n-2}, \nabla^2 f_n$			
…	$\triangle f_{n-1}, \nabla f_n$				
f_n					

性质：1°　$\triangle^m f_i = \nabla^m f_{i+m}$

2°　$f[x_0, x_1, \cdots, x_m] = \dfrac{1}{m!\ h^m} \triangle^m f_0 = \dfrac{1}{m!\ h^m} \nabla^m f_m$

利用性质 2°将牛顿插多项式(3.16)中的差商用差分代替，得到：

$$P_n(x) = P_n(x_0 + th) = f_0 + \triangle f_0 t + \frac{\triangle^2 f_0}{2!} t(t-1) + \cdots$$

$$+ \frac{\triangle^m f_0}{n!} t(t-1)\cdots(t-n+1) \tag{3.19}$$

式(3.19)叫牛顿向前插值多项式，其系数由差分表中的上数第一行的差分决定.

同理由牛顿插值多项式(3.18)得到：

$$P_n(x) = P_n(x_n + th) = f_n + \nabla f_n t + \frac{\nabla^2 f_n}{2!} t(t+1) + \cdots$$

$$+ \frac{\nabla^n f_n}{n!} t(t+1)\cdots(t+n-1) \tag{3.20}$$

式(3.20)叫牛顿向后插值多项式，其系数由差分表中倒数第一行的差分决定.

上面公式的用法：当我们对x_0附近的x进行插值时，用前插公式；倘若要对x_n附近的x进行插值，自然用后插公式.

(3.19)的余项为：

$$R_n(x) = \frac{h^{n+1}}{(n+1)!} f^{(n+1)}(\xi) t(t-1)\cdots(t-n), \xi \in [a,b] \tag{3.21}$$

式(3.20)的余项为：

$$R_n(x) = \frac{h^{n+1}}{(n+1)!} f^{(n+1)}(\xi) t(t+1)\cdots(t+n-1), \xi \in [a,b] \tag{3.22}$$

三、氨蒸汽的压力和密度

在化学工程中经常会遇到高温状态下蒸汽的压力和密度的问题，但是考虑到测量设备等的限制，我们希望利用低温状态下的压力等有关数据进行外推，下表给出了氨蒸汽的一组温度和压力数据，并设氨的潜热等于 1 265kJ/kg.

表 3-7

温度	℃	20	25	30	35	40	45	50	55	60
压力	(kN/m^2)	805	985	1 170	1 365	1 570	1 790	2 030	2 300	2 610

试从已知数据中计算出 75℃氨蒸汽的压力和密度.

解:首先从表 3-7 的数据可知结点为等距,所以可制成下面的差分表.

表 3-8

t — P	$\nabla P,\Delta P$	$\nabla^2 P,\Delta^2 P$	$\nabla^3 P,\Delta^3 P$	$\nabla^4 P,\Delta^4 P$
20——805				
	180			
25——985		5		
	185		5	
30——1 170		10		5
	195		0	
35——1 365		10		−5
	205		5	
40——1 570		15		0
	210		5	
45——1 790		20		5
	240		10	
50——2 030		30		0
	270		10	
55——2 300		40		
	310			
60——2 610				

表中的四阶差分在±5 之间摆动,因此 $P=f(t)$可用三次多项式近似代替,也就是说用到三阶差分为止. 由于 $t=75$℃,在 60℃附近,所以取 60℃为基准温度,于是 $P=P(t)=P(t_m+mh)$

式中 $t=75$℃,$t_m=60$℃,$h=5$℃,$m=3$

利用表(3-8)中最后一行差分,及式(3.20)得

$$P=P_n+\nabla P_n\cdot m+\nabla^2 P_n\cdot\frac{m(m+1)}{2!}+\nabla^2 P_n\cdot\frac{m(m+1)(m+2)}{3!}$$

$$=2\,610+310\times 3+40\times\frac{3\times 4}{2!}+10\times\frac{3\times 4\times 5}{3!}=3880\text{kN/m}^2$$

从而求得 75℃氨蒸汽的压力为 3 880kN/m²

其次,为了求得氨蒸汽的密度,利用(Clausius—Clapeyron)方程

$$\frac{dP}{dT}=\frac{\delta H}{T(V_V-V_L)}$$

其中 T 是 k 氏温度(K),V_V,V_L 分别表示蒸汽和液体的体积,δH 是氨的潜热. 假设液体体积和蒸汽体积相比可以忽略,则由方程可知氨的密度 ρ 满足

$$\rho=\frac{1}{V_V}=\frac{T}{\delta H}\cdot\frac{dP}{dT}$$

只要求出$\frac{dP}{dT}$的值,即可由上式求得 75℃时的氨的密度.

由于用三次多项式

$$P(t)=P(60+5m)=P_n+\nabla P_n m+\nabla^2 P_n\frac{m(m+1)}{2!}+\nabla^3 P_n\frac{m(m+1)(m+2)}{3!}$$

代替 P 的合理性；$\frac{dP}{dT}$可由数值微分，即对插值多项式 $P(t)$ 求导得到. 而产生的误差为：

$$\left[\frac{h^4}{4!}f^{(4)}(\xi)m(m+1)(m+2)(m+3)\right]_t'.$$

由于四阶差分在 0 的附近摆动，因此这个误差中 $f^{(4)}(\xi)$ 也在 0 的附近摆动.

而
$$\frac{dP}{dt}=\frac{dP(60+5m)}{dm}\cdot\frac{dm}{dt}=\frac{1}{5}\frac{dP(60+50m)}{dm}\triangleq\frac{1}{5}\frac{dP}{dm}$$

$$\frac{dP}{dm}=\nabla P_n+\frac{\nabla^2 P_n}{2!}(2m+1)+\frac{\nabla^3 P_n}{3!}(3m^2+6m+2)$$

∴
$$\frac{dP}{dT}=\frac{dP}{dt}\bigg|_{t=75^\circ}=\frac{1}{5}\frac{dP}{dm}\bigg|_{m=3}=\frac{1}{5}(310+140+\frac{470}{3})$$

从而
$$\rho=\frac{273+75}{1265}\cdot\frac{1}{5}(310+140+\frac{470}{3})=29.0\text{kg/m}^3$$

注：我们之所以能对 75℃氨的压力密度作出合理的计算，是因为这个温度值与测量的温度之间差别不太大，可以外推求得. 另外此问题也可采用牛顿前插多项式，利用向前差分，取基准温度为 $t_0=45$℃，读者不妨试着做一下.

第三节　三次样条插值及模型

一、三次样条插值

前面我们介绍了构造插值多项式的方法，从直观想象，多项式次数越高，逼近的精度会越好. 而实际则不然，一般很少采用七八次以上的插值多项式. 因为高次插值的稳定性得不到保证. 当插值结点较多时，可采用分段低次多项式插值.

这种方法大致分为两类.

一类是局部化分段插值. 即把插值区间分段后，在每一子段上直接构造低次多项式，如分段线性插值，分段抛物插值等.

另一类是非局部化分段插值，如样条插值. 因为简单分段插值光滑性较差，满足不了某些工程技术实际问题的要求. 例如，在船体、飞机等外形曲线的设计中，不仅要求曲线连续而且还要求曲线的曲率连续. 为解决这一类型问题，就产生了三次样条插值.

1. 三次样条函数的定义

给定 $n+1$ 个结点，$x_0<x_1<\cdots<x_{n-1}<x_n$，及对应函数 $y=f(x)$ 的值，y_i，$i=0,\cdots,n$. 如果$[x_0,x_n]$上的分段函数 $s(x)$ 满足：

(1). 在每个子区间$[x_i,x_{i+1}]$上的表达式 $s_i(x)$ 是次数不高于 3 的代数多项式；

(2). 在每个内结点 $x_i(i=1,2,\cdots,n-1)$ 上具有二阶连续导数；

(3). $s(x_i)=y_i$，$i=0,1,\cdots,n$；

则称 $s(x)$ 是三次样条插值函数，简称三次样条.

几何上，样条曲线是由分段三次曲线连接而成，在连接点上，不仅函数自身是连续的，

而且它的一阶和二阶导数也是连续的.

2. 三次样条插值的边界条件

构造三次样条函数,除用 $n+1$ 个结点的函数值以外,还需补充两个条件,通常在区间 $[x_0,x_n]$ 的端点各加一个条件,称为边界条件.

常见的边界条件有下列三种.

第一类边界条件:$s'(x_0),s'(x_n)$ 取定值.

第二类边界条件:$s''(x_0),s''(x_n)$ 取定值.

特别,边界条件 $s''(x_0)=s''(x_n)=0$ 称为自然边界条件.

第三类边界条件:$s(x_0)=s(x_n),s'(x_0)=s'(x_n),s''(x_0)=s''(x_n)$,称为周期边界条件.

3. 求样条插值函数的三转角法

用结点上的一阶导数值 $s'(x_i)=m_i(i=0,1,\cdots,m)$ 作参数表示 $s(x)$,通过确定 m_i 来求出 $s(x)$ 的方法称为三转角法.

①首先求三次埃尔米特插值多项式

设已知函数 $f(x)$ 在 x_i,x_{i+1} 两互异点的函数值及其导数值.

$$\begin{cases} f(x_i)=y_i, \\ f'(x_i)=m_i, \end{cases} \qquad \begin{cases} f(x_{i+1})=y_{i+1}, \\ f'(x_{i+1})=m_{i+1}. \end{cases}$$

求三次多项式 $P_i(x)$,使其满足

$$\begin{cases} P_i(x_i)=y_i, \\ P'_i(x_i)=m_i, \end{cases} \qquad \begin{cases} P_i(x_{i+1})=y_{i+1}, \\ P'_i(x_{i+1})=m_{i+1}. \end{cases}$$

这样的 $P_i(x)$ 称为三次埃尔米特插值多项式.

解决上述问题,仍采用基函数的思想,求三次多项式 $\varphi_i(x)$ 和 $\varphi_{i+1}(x)$,使它们分别满足下述条件

$$\begin{cases} \varphi_i(x_i)=y_i, \\ \varphi'_i(x_i)=m_i, \\ \varphi_i(x_{i+1})=0, \\ \varphi'_i(x_{i+1})=0; \end{cases} \qquad \begin{cases} \varphi_{i+1}(x_i)=0, \\ \varphi'_{i+1}(x_i)=0, \\ \varphi_{i+1}(x_{i+1})=y_{i+1}, \\ \varphi'_{i+1}(x_{i+1})=m_{i+1}. \end{cases}$$

由 $\varphi_i(x)$ 是三次多项式及在 x_{i+1} 点所满足的条件,x_{i+1} 是它的二重点,即它有因子 $(x_{i+1}-x)^2$,记 $h_i=x_{i+1}-x_i$,则 $\varphi_i(x)$ 可写成

$$\varphi_i(x) = y_i[2(x-x_i)+h_i](x_{i+1}-x)^2/h_i^3 + m_i(x-x_i)(x_{i+1}-x)^2/h_i^2 \tag{3.23}$$

同理

$$\varphi_{i+1}(x) = y_{i+1}[2(x_{i+1}-x)+h_i](x-x_i)^2/h_i^3 - m_{i+1}(x_{i+1}-x)(x-x_i)^2/h_i^2 \tag{3.24}$$

则
$$P_i(x) = \varphi_i(x)+\varphi_{i+1}(x), \qquad i=0,1,\cdots,n-1. \tag{3.25}$$

(2)其次利用三次埃尔米特插值多项式公式(3.25)构造分段三次多项式

$$s_i(x) = \varphi_i(x)+\varphi_{i+1}(x), x\in[x_i,x_{i+1}], (i=0,1,\cdots,n-1) \tag{3.26}$$

这样建立的 $s(x)$ 已满足 $s(x_i-0)=s(x_i+0),s'(x_i-0)=s'(x_i+0)(i=1,2,\cdots,n-1)$,

但是，$s''(x)$在内结点处未必连续，为此令

$$s''(x+0)=s''(x_i-0),(i=1,2,\cdots,n-1)$$

得方程组为：

$$\lambda_i m_{i-1}+2m_i+(1-\lambda_i)m_{i+1}=g_i \quad (i=1,2,\cdots,n-1) \tag{3.27}$$

其中

$$\begin{cases}\lambda_i=\dfrac{h_i}{h_{i-1}+h_i},\mu_i=1-\lambda_i,\\ g_i=3\left(\lambda_i\dfrac{y_i-y_{i-1}}{h_{i-1}}+\mu_i\dfrac{y_{i+1}-y_i}{h_i}\right).\end{cases} \tag{3.28}$$

注意：λ_i,u_i,g_i 都可事先计算出来，故式(3.27)是关于 $n+1$ 个未知量 $m_0,m_1,\cdots m_n$ 的 $n-1$ 个线性方程，若要唯一确定 m_i，还需要利用边界条件补充两个方程，然后利用追赶法求解方程组，从而确定 $m_i(i=1,2,\cdots,n-1)$，把 m_i 代入式(3.26)即可求得三次样条插值函数.

1° 对第一类边界条件，由于 $m_0=f'(x_0)$，$m_n=f'(x_n)$已知，方程组(3.27)实为只含 $n-1$ 个未知量 $m_1,m_2,\cdots,m_{n-1}$的 $n-1$ 个方程，显然这个方程组可唯一确定，$m_1,m_2,\cdots,m_{n-1}$，从而得到 $s(x)$.

2° 对第二类边界条件，由于 $M_0=f''(x_0)$，$M_n=f''(x_n)$已知可得

$$\begin{cases}2m+m_1=g_0,\\ m_{n-1}+2m_n=g_n.\end{cases} \tag{3-29}$$

其中 $$g_0=3\frac{y_1-y_0}{h_0}-\frac{h_0}{2}M_0,\quad g_0=3\frac{y_n-y_{n-1}}{h_{n-1}}+\frac{h_{n-1}}{2}M_0$$

3° 对周期边界条件，有

$$\begin{cases}m_0=m_n,\\ \mu_n m_1+\lambda_n m_{n-1}+2m_n=g_n.\end{cases} \tag{3.30}$$

其中

$$\begin{cases}\lambda_n=\dfrac{h_0}{h_{n-1}+h_0},\quad \mu_n=1-\lambda_n,\\ g_n=3\left(\mu_n\dfrac{y_1-y_0}{h_0}+\lambda_n\dfrac{y_n-y_{n-1}}{h_{n-1}}\right).\end{cases}$$

例 1 给定函数 $y=f(x)$的数据表：

表 3-9

x_i	1	2	4	5
y_i	1	3	4	2
y'_i	$\frac{17}{8}$			$-\frac{19}{8}$

试求 $y=f(x)$的三次样条函数 $s(x)$，并求 $f(3)$和 $f(4,5)$的近似值.

解：由 $x_0=1,x_1=2,x_2=4,x_3=5$ 得：$h_0=1,h_1=2,h_2=1$.

由公式(3.28)得 $\lambda_1=\frac{2}{3},\lambda_2=\frac{1}{3},\mu_1=\frac{1}{3},\mu_2=\frac{2}{3},g_1=\frac{9}{2},g_2=-\frac{7}{2}$.

此时方程组(3.27)为

$$\begin{cases} 2m_1 + \mu_1 m_2 = g_2 - \lambda_1 m_0, \\ \lambda_1 m_1 + 2m_2 = g_2 - \mu_2 m_3. \end{cases}$$

注意 $m_0 = \frac{17}{8}, m_3 = -\frac{19}{8}$,则上述方程组为

$$\begin{cases} 2m_1 + \frac{1}{3}m_2 = \frac{37}{12}, \\ \frac{1}{3}m_1 + 2m_2 = -\frac{23}{12}. \end{cases}$$

解此方程得 $m_1 = \frac{7}{4}, m_2 = -\frac{5}{4}$.

把 m_0, m_1, m_2, m_3 代入(3.26)知 $s(x)$ 在 $[x_0, x_1] = [1,2]$ 上的表达式为

$$\begin{aligned} s(x) &= (2x-1)(2-x)^2 + 3(5-2x)(x-1)^2 \\ &\quad + \frac{17}{8}(x-1)(2-x)^2 - \frac{7}{4}(2-x)(x-1)^2 \\ &= -\frac{1}{8}x^3 + \frac{3}{8}x^2 + \frac{7}{4}x - 1, \qquad x \in [1,2] \end{aligned}$$

同理可得 $s(x)$ 在其它子区间上的表达式为

$$s(x) = -\frac{1}{8}x^3 + \frac{3}{8}x^2 + \frac{7}{4}x - 1, \qquad x \in [2,4]$$

$$s(x) = \frac{3}{8}x^3 - \frac{45}{8}x^2 + \frac{103}{4}x - 33, \qquad x \in [4,5]$$

故所求的三次样条为

$$s(x) = \begin{cases} -\frac{1}{8}x^3 + \frac{3}{8}x^2 + \frac{7}{4}x - 1, & x \in [1,4], \\ \frac{3}{8}x^3 - \frac{45}{8}x^2 + \frac{103}{4}x - 33, & x \in [4,5]. \end{cases}$$

据此得:

$$f(3) \approx s(3) = \frac{17}{4},$$

$$f(4.5) \approx s(4.5) = \frac{201}{64}.$$

二、动物上半身曲线的描绘

问题:用三次样条插值描绘图 3-2 中大家熟知图形的上半部分曲线.



图 3-2

解：首先建立坐标系如图 3-3 所示，然后测出曲线型值点数据（表 3-10），其中插值结点 x_i 选取在曲线变化比较剧烈的地方．因为这个曲线有两个导数不连续点（头与爪子，头与脊背交界处），所以分段使用三次样条插值．

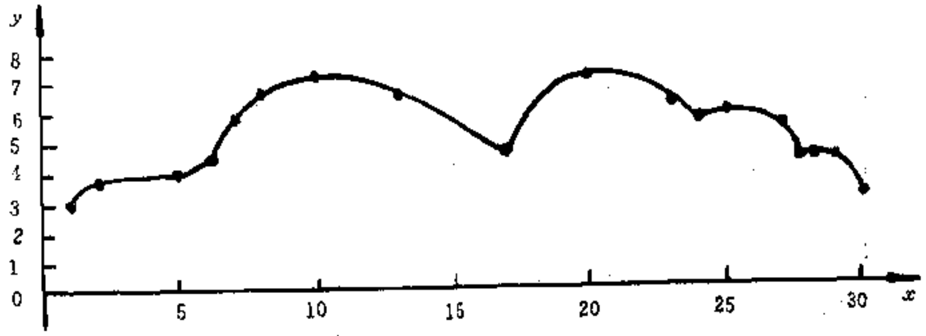


图 3-3

表 3-10

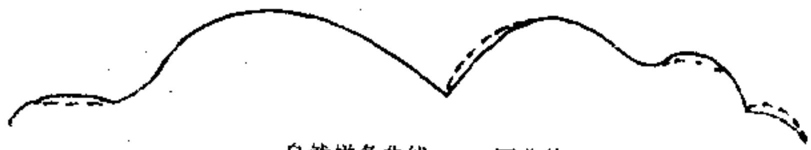
曲线 1			曲线 2			曲线 3		
i	x_i	$f(x_i)$	i	x_i	$f(x_i)$	i	x_i	$f(x_i)$
0	1	3.0	0	17	4.5	0	27.7	4.1
1	2	3.7	1	20	7.0	1	28	4.3
2	5	3.9	2	23	6.1	2	29	4.1
3	6	4.2	3	24	5.6	3	30	3.0
4	7	5.7	4	25	5.8			
5	8	6.6	5	27	5.2			
6	10	7.1	6	27.7	4.1			
7	13	6.7						
8	17	4.5						

方案 I　用三次自然样条插值计算．

利用边界条件是自然边界条件(3.29)，分别计算出三段曲线型值点处的参数 m_i（表 3-11）然用由公式(3.26)可得自然样条的表达式，其图形见图 3-4．

表 3-11

曲线 1			曲线 2			曲线 3		
i	x_i	m_i	i	x_i	m_i	i	x_i	m_i
0	1	0.786	0	17	1.106	0	27.7	0.749
1	2	0.529	1	20	0.289	1	28	0.503
2	5	−0.086	2	23	−0.660	2	29	−0.787
3	6	1.019	3	24	−0.137	3	30	−1.256
4	7	1.408	4	25	0.306			
5	8	0.547	5	27	−1.263			
6	10	0.049	6	27.7	−1.725			
7	13	−0.342						
8	17	−0.63						



自然样条曲线——，原曲线----

图 3-4

方案 I　用第一类边界条件确定的样条插值.

首先测得每段曲线端点的斜率，见表 3-12，然后利用第一类边界条件可计算出每段曲线其它型值点处参数的 m_i 的值（表 3-12），然后，利用公式(3.26)可得 I 型样条函数，其图形见图 3-5.



I 型样条曲线——，原曲线----

图 3-5

表 3-12

曲线 1				曲线 2				曲线 3			
i	x_i	m_i	备注	i	x_i	m_i	备注	i	x_i	m_i	备注
0	1	1.00	测量值	0	17	3.00	测量值	0	27.7	0.33	测量值
1	2	0.447	计算值	1	20	−0.189	计算值	1	28	0.661	计算值
2	5	−0.074	计算值	2	23	−0.609	计算值	2	29	−0.765	计算值
3	6	1.016	计算值	3	24	−0.111	计算值	3	30	−1.5	测量值
4	7	1.409	计算值	4	25	0.154	计算值				
5	8	0.547	计算值	5	27	0.401	计算值				
6	10	0.048	计算值	6	27.7	−4.00	测量值				
7	13	−0.339	计算值								
8	17	−0.67	测量值								

三、用自然三次样条预测人口的数值实验

下表是我国 1949 至 1984 年间的每隔 5 年的人口数，采用自然样条插值计算出 1949 至 1984 年间每二年的人口数，并由此预测我国 1990 年的人口数.

x_i(年)	1949	1954	1959	1964	1969	1974	1979	1984
y_i(亿人)	5.416 7	6.026 6	6.720 9	7.049 9	8.067 1	9.085 9	9.754 2	10.347 5

数值实验的程序

```
100  REM  三次样条插值
110  CLS
120  READ N:N=N-1
130  DIM X(N),Y(N),H(N),F(N),A(N),U(N),G(N),B(N),L(N),D(N),M
     (N)
140  FOR J=0 TO N:READ X(J),Y(J):NEXT J
150  FOR J=0 TO N-1
160  H(J)=X(J+1)-X(J):F(J)=(Y(J+1)-Y(J))/H(J)
170  NEXT J
180  FOR J=1 TO N-1
190  A(J)=H(J)/(H(J-1)+H(J))
200  U(J)=H(J-1)/(H(J-1)+H(J))
210  G(J)=3*(A(J)*F(J-1)+U(J)*F(J))
220  NEXT J
230  G(0)=3*F(0):G(N)=3*F(N-1)
240  FOR I=0 TO N:B(I)=2:NEXT I
250  U(0)=1
260  L(0)=U(0)/B(0):D(0)=G(0)/B(0)
270  FOR K=1 TO N-1
280  L(K)=U(K)/(B(K)-A(K)*L(K-1))
290  D(K)=(G(K)-A(K)*D(K-1)/(B(K)-A(K)*L(K-1))
300  NEXT K
310  M(N)=(G(N)-A(N)*D(N-1))/(B(N)-A(N)*L(N-1))
320  FOR K=N-1 TO 0 STEP-1
330  M(K)=D(K)-L(K)*M(K+1)
340  NEXT K
350  PRINT "请输入插值数据:";PRINT
360  INTPUT "X=";X
370  FOR J=0 TO N-1
380  IF X<X(0) THEN 420
390  IF X>X(N) THEN J=N-1:GOTO 420
400  IF X>=X(J) AND X<=X(J+1) THEN 420
410  NEXT J
420  S1=(X-X(J+1))^2*(H(J)+2*(X-X(J)))/(H(J)^3)*Y(J)
430  S2=(X-X(J))^2*(H(J)+2*(X(J+1)-X))/(H(J)^3)*Y(J+1)
440  S3=(X-X(J+1))^2*(X-X(J))/(H(J)^2)*M(J)
450  S4=(X-X(J))^2*(X-X(J+1))/(H(J)^2)*M(J+1)
460  SX=S1+S2+S3+S4
```

```
470  Y0=CSRLIN:LOCATE Y0-1,20:PRINT"y=";SX:GOTO 360
480  END
490  DATA 8
500  DATA 1 949,5.416 7,1954,6.026 6,1 959,6.720 9,1 964,7.049 9,1 969,8.
     067 1,8.067 1,1 974,9.085 9,1 979,9.754 2,1984,10.347 5
```

请输入插值数据：

x=1 951	y=5.639937
x=1 953	y=5.886 858
x=1 956	y=6.335 171
x=1 958	y=6.616 942
x=1 961	y=6.840 153
x=1 963	y=6.951 258
x=1 966	y=7.384 36
x=1 968	y=7.832 064
x=1 971	y=8.511 841
x=1 973	y=8.908 771
x=1 976	y=9.393 909
x=1 978	y=9.646 25
x=1 981	y=9.951 34
x=1 983	y=10.188 27
x=1 990	y=12.524 98

第四节　双线性插值及模型

一、双线性插值

双线性插值是二维插值法中最简单的一种，是分段线性插值的推广，这一部分不介绍二次或二次以上的二维插值，但可以通过了解双线性插值后，也很自然把前面学过的一维的插值法，如三次样条插值推广到二维的双三次样条.

给定矩形网络 $\Delta=\Delta x\times\Delta y$

$\Delta x: a=x_0<x_1<\cdots<x_N=b,$

$\Delta y: c=y_0<y_1<\cdots<y_M=d,$

以上函数值

$$f(x_i,y_i)=f_{ij}\quad(0\leqslant i\leqslant N,0\leqslant j\leqslant M)$$

分片双线性插值函数为：

$$S_1(x,y)=\sum_{i=0}^{N}\sum_{j=0}^{M}f_{ij}\varphi_{ij}(x,y)\tag{3.31}$$

其中 $\varphi_{ij}(x,y)$是双线性插值的基函数.

双线性插值的基函数形式为(见图 3-6)

$$
\varphi_{ij}(x,y)=\begin{cases}\dfrac{x-x_{i+1}}{x_i-x_{i+1}}\cdot\dfrac{y-y_{j+1}}{y_j-y_{j+1}}, & \text{区域 I};\\[2mm] \dfrac{x-x_{i-1}}{x_i-x_{i-1}}\cdot\dfrac{y-y_{j+1}}{y_j-y_{j+1}}, & \text{区域 II};\\[2mm] \dfrac{x-x_{i-1}}{x_i-x_{i-1}}\cdot\dfrac{y-y_{j-1}}{y_j-y_{j-1}}, & \text{区域 III};\\[2mm] \dfrac{x-x_{i+1}}{x_i-x_{i+1}}\cdot\dfrac{y-y_{j-1}}{y_j-y_{j-1}}, & \text{区域 IV};\\[2mm] 0, & \text{其它}.\end{cases} \tag{3.32}
$$

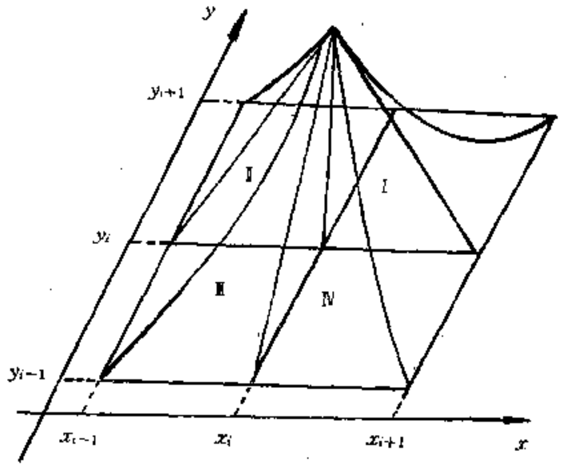


图 3-6

即 $\varphi_{ij}(x,y)$为两个方向线性基函数的乘积:

$$\varphi_{ij}=L_i(x)L_j(y).$$

二、三维空间雷达反射因子的一种求解方法

人们在利用雷达探测天气时,由于受时限的影响,往往只能获取不超过 20 个锥面上的云层的反射因子(利用反射因子可以提取气象因素).但是为了进行合理的分析和显示,又必须知道最好是某个空间范围内的较密网格上的数据,这就给我们提出了这样的一个问题,能否利用空间中有限个锥面上的数据生成所需的数据网呢?

回答是肯定的.首先我们建立以雷达站为中心的球坐标系,取 z 轴的正向为铅直向上的方向.这里我们之所以选球坐标系而没有选用直角坐标系,是因为 φ 固定时表示一个锥面,能体现出该模型的优越性.分别用 φ、θ 和 R 表示仰角、方位角和极径(空间点到雷达站的距离).用 z_e 表达雷达探测到的反射因子.因为雷达并不是连续采样的,所以我们所得到的反射因子是由球坐标给出的空间网格点上的值.设雷达采样的网格点为$\{(R_i,\theta_j,\varphi_k):i=1,z,\cdots,m;j=1,2,\cdots,n;k=1,2,\cdots,l\}$即已知 z_e 在上述网格点上的数据:$z_e(R_i,\theta_j,\varphi_k)$,为简便起见,记为 $z_e(i,j,k)$.

下面我们希望利用 $z_e(i,j,k)$的值计算空间任一点(R,θ,φ)的反射因子 $z_e(R,\theta,\varphi)$.很显然,该问题属于三维插值法,我们可以把该问题分解为一个二维插值(双线性插值)和一个一维插值(分段线性插值)两步骤来实现.

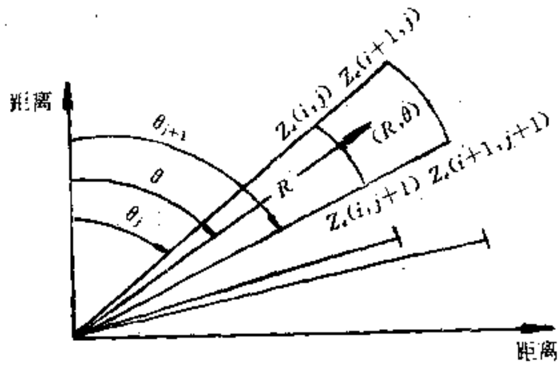


图 3-7

第一步:设插值点(R,θ,φ)当 φ 固定时在(R,θ)平面上落在由 $R=R_j,R=R_{j+1}$及 $\theta=\theta_i,\theta=\theta_{i+1}$所围成的区域内(见图 3-7),我们利用关于极径 R 和方位角 θ 的双线性插值来计算 $z_e(R,\theta,\varphi)$,因为这时 φ 是固定的,所以简记成 $z_e(R,\theta)$.即用四个邻近的反射因子值 $z_e(i,$

$j)$，$z_e(i,j+1)$，$z_e(i+1,j)$，$z_e(i+1,j+1)$求 $z_e(R,\theta)$. 计算公式如下：

$$z_e(R,\theta)=\frac{R-R_i}{R_{i+1}-R_i}\cdot\frac{\theta-\theta_j}{\theta_{j+1}-\theta_j}\cdot z_e(i+1,j+1)+\frac{R-R_i}{R_{i+1}-R_i}\cdot\frac{\theta-\theta_{j+1}}{\theta_j-\theta_{j+1}}z_e(i+1,j)$$
$$+\frac{R-R_{i+1}}{R_i-R_{i+1}}\cdot\frac{\theta-\theta_{j+1}}{\theta_j-\theta_{j+1}}z_e(i,j)+\frac{R-R_{i+1}}{R_i-R_{i+1}}\cdot\frac{\theta-\theta_j}{\theta_{j+1}-\theta_j}z_e(i,j+1)$$

第二步：为了得到插值点(R,θ,φ)的反射因子值，最后还需要用前面得到的点(R,θ,φ_k)和(R,θ,φ_{k+1})进行线性插值，这里假设 φ 落在区间$(\varphi_k,\varphi_{k+1})$内，其公式如下：

$$z_e(R,\theta,\varphi)=\frac{\varphi-\varphi_k}{\varphi_{k+1}-\varphi_k}\cdot z_e(R,\theta,\varphi_{k+1})+\frac{\varphi-\varphi_{k+1}}{\varphi_k-\varphi_{k+1}}z_e(R,\theta,\varphi_k)$$

这第二步插值如图 3-8 所示. 至此，所讨论的空间范围内任一点(R,θ,φ)的反射因子均可以利用上述两公式求出，从而能够显示和分析.

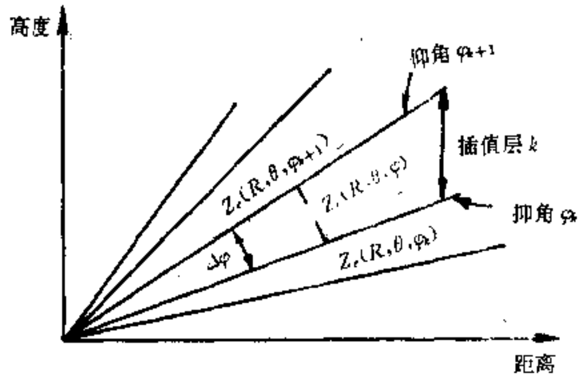


图 3-8

第五节　曲线拟合法及其模型

设通过实验等方法测得反映某个一元函数 $y=f(x)$的数据$(x_i,y_i)(i=0,1,\cdots,m)$，现要求利用这些数据构造出 $f(x)$的近似表达式 $p(x)$. 但由于实验观测到的数据不可避免的带有误差，因此并不要求曲线 $y=p(x)$(图 3-9)经过每一个数据点$(x_i,y_i)(i=0,1,\cdots,m)$，而且是要求它尽可能地从测定点的附近通过，而只要求其偏差 $r_i=p(x_i)-y_i(i=0,1,\cdots,m)$的某种度量取最小，这就是曲线拟合问题. 这样的函数 $p(x)$称为拟合函数，

本节主要介绍曲线拟合中最常用的最小二乘法的原理，在用最小二乘法求拟合函数的过程中，先根据讨论问题的特点及经验，确定 $p(x)$的类型，然后确定 $p(x)$中待定的参数.

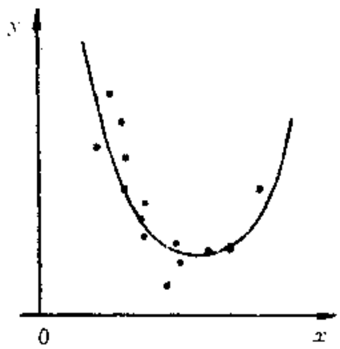


图 3-9

一、最小二乘原理和多项式拟合

对给定的数据$(x_i,y_i)(i=0,1,\cdots,m)$，求 $p(x)$，使偏差 $r_i=p(x_i)-y_i$ 的平方和最小，即按使

$$\sum_{i=0}^{m}[p(x_i)-y_i]^2$$

取最小的原则确定 $p(x)$，称这种方法为最小二乘法.

若选择 $p(x)$是次数为 $n(n\leqslant m)$的多项式，这样的曲线拟合问题叫作多项式拟合.

设 $$p(x) \xlongequal{\triangle} p_n(x) = a_0 + a_1 x + \cdots + a_n x^n = \sum_{k=0}^{n} a_k x^k$$

使其满足

$$F(a_0, a_1, \cdots, a_n) = \sum_{i=0}^{m} [p_n(x_i) - y_i]^2 = \sum_{i=0}^{m} [\sum_{k=0}^{n} a_k x_i^k - y_i]^2$$

最小，由于 F 可以看作是 $a_0, a_1, \cdots, a_n$ 的多元函数，所以上述问题为多元函数极值问题. 由多元函数极值的必要条件知，$a_k(k=0,1,\cdots,n)$ 满足

$$\frac{\partial F}{\partial a_0} = 0, \frac{\partial F}{\partial a_1} = 0, \cdots, \frac{\partial F}{\partial a_n} = 0.$$

于是得到关于 $a_0, a_1, \cdots, a_n$ 的线性方程组为

$$\begin{pmatrix} m+1 & \sum_{i=0}^{m} x_i & \cdots & \sum_{i=0}^{m} x_i^n \\ \sum_{i=0}^{m} x_i & \sum_{i=0}^{m} x_i^2 & \cdots & \sum_{i=0}^{m} x_i^{n+1} \\ \cdots & \cdots & \cdots & \cdots \\ \sum_{i=0}^{m} x_i^n & \sum_{i=0}^{m} x_i^{n+1} & \cdots & \sum_{i=0}^{m} x_i^{2n} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} \sum_{i=0}^{m} y_i \\ \sum_{i=0}^{m} x_i y_i \\ \vdots \\ \sum_{i=0}^{m} x_i^n y_i \end{pmatrix} \tag{3.32}$$

称为法方程组或正则方程组.

法方程组的系数矩阵为对称阵，若它非奇异可唯一确定 $p_n(x)$.

例 1 测得铜导线在温度 t_i 时的电阻 r_i 如表 3-13，求电阻 r 与温度 t 的近似函数关系式，并求 $r=0$ 时的温度.

表 3-13

i	0	1	2	3	4	5	6
t_i	19.1	25.0	30.1	36.0	40.0	45.1	50.1
r_i	76.30	77.80	79.25	80.80	82.35	83.90	85.10

解：根据所给数据，在坐标纸上描点(图 3-10)，可以看出点的分布形状近似地为一条直线所以取一次式

$$r = a + bt$$

为拟合函数，因为

$$\sum_{i=0}^{6} t_i = 246.3 \qquad \sum_{i=0}^{6} t_i^2 = 9\ 325.83$$

$$\sum_{i=0}^{6} r_i = 565.3 \qquad \sum_{i=0}^{6} r_i t_i = 20\ 029.445$$

所以法方程组为

$$\begin{pmatrix} 7 & 245.3 \\ 245.3 & 9\ 325.83 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 565.3 \\ 20\ 029.445 \end{pmatrix}$$

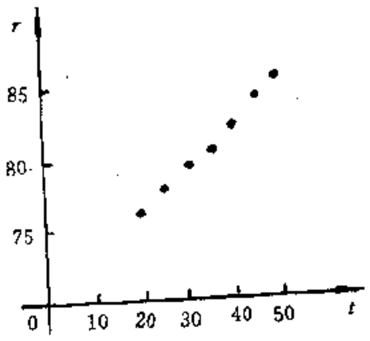


图 3-10

解之得

$$a = 70.572, \quad b = 0.291$$

所以得 r 与 t 的拟合直线为：

$$r = 70.572 + 0.291t$$

并预测得 $r=0$ 时，$t=-242.5$，即预测温度为 -242.5℃时，铜导线无电阻.

注意：用最小二乘法求拟合曲线时，首先要确定 $P(x)$ 的形式，这不单是数学问题，还与所研究问题的运动规律及所得观测数据 (x_i, y_i) 有关，通常是先用给定数据描图，确定 $P(x)$ 的类型，这种确定并不是一次就能完成. 往往要经过确定几个类型后，通过实际计算比较均方误差

$$\sqrt{\sum_{i=0}^{m}[y_i - p(x_i)]^2}$$

选择均方误差较小的拟合函数，一般说来拟合函数中待定系数越少越好.

例 2　某日每隔一小时，测量一次气温的数据如下表

表 3-14

时间 t	0	1	2	3	4	5	6	7	8	9	10	11	12
温度 c	15	14	14	14	14	15	16	18	20	22	23	25	28
时间 t	13	14	15	16	17	18	19	20	21	22	23	24	
温度 c	31	32	31	29	27	25	24	22	20	18	17	16	

试用最小二乘法描述这天温度变化规律的近似函数.

解：从气温表 3-14 看出，在 0～5 时之间有最低气温，在 13～15 时之内有最高气温，所以用三次多项式来拟合是合理的. 下面取 $x_i = t_i - 12$，$y_i = c(t_i) - 14$，令

$$f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3$$

并利用 Mathematica 软件包计算结果.

```
ln[1]:=e={{-12,1},{-11,0},{-10,0},{-9,0},{-8,0},{-7,1},{-6,2},
    {-5,4},{-4,6},{-3,8},{-2,9},{-1,11},{0,14},{1,17},{2,18},
    {3,17},{4,15},{5,13},{6,11},{7,10},{8,8},{9,6},{10,4},{11,3},
    {12,2}};
ln[2]:=x=.;f[x]=Fit[e,{1,x,x^2,x^3},x]
out[2]=12.067 6+1.091 5x-0.093 609 8x²-0.007 963 69x³
ln[3]:=g=ListPlot[e];
ln[4]:=h=Plot[f[x],{x,-12,12}];
ln[5]:=Show[g,h]
out[5]=-Graphics-
ln[6]:=Table[f[x],{x,-12,12,1}]
out[6]={-0.748 954,-0.666 014,-0.244 689,0.467 237,1.421 98,2.571 77,
    3.868 81,5.265 3,6.713 52,8.165 63,9.573 87,10.890 5,12.067 6,
    13.057 5,13.812 5,14.284 6,14.426 2,14.189 4,13.526 5,12.389 7,
```

10.731 2,8.503 19,5.657 94,2.147 66,−2.075 45}

从而得拟合函数 $c(t)=14+f(t-12)$ 即

$$c(t)=26.0676+1.0915(t-12)-0.0936098(t-12)^2-0.0079369(t-12)^3$$

图 3-11 是数据点与拟合曲线的叠加图,其均方误差由 out[6]的数据得

$$\sqrt{\sum_{i=0}^{24}[y_i-f(x_i)]^2}=10.29942$$

从图 3-11 中看到用三次多项式拟合原数据效果并不理想,而且均方误差也较大,为得到更理想的拟合效果,采用增加拟合多项式次数的方法,当然拟合函数与叠加图也都是由 Mathematiea 软件包上机得到.

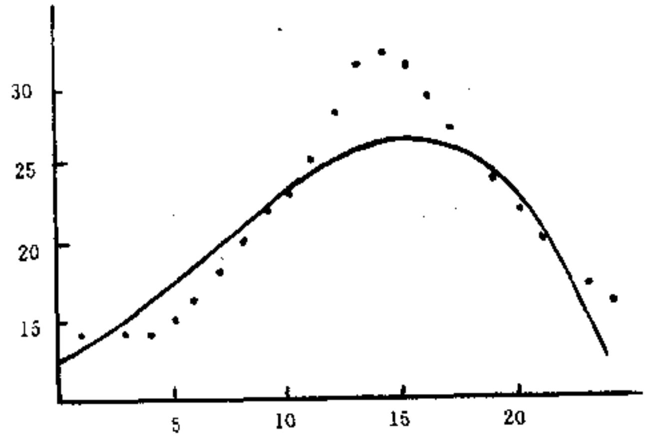


图 3-11 (三次拟合)

利用五次多项式拟合原数据得,拟合函数为

$$c(t)=15.0538-0.830413x+0.0156656x^2+0.0450041x^3-0.00367513x^4+0.00007659x^5$$

从图 3-12 中可以看到拟合的效果要好于三次多项式拟合.

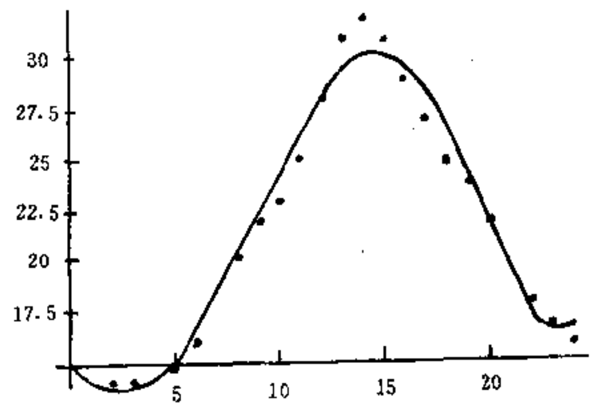


图 3-12 (五次拟合)

利用七次多项式拟合原数据,拟合函数为

$$c(t)=15.0304-1.74752x+0.905481x^2-0.225824x^3+0.0331076x^4-0.00240188x^5+0.0000810847x^6-0.00000102721x^7$$

从图 3-13 中可看到用七次多项式拟合的效果是很理想的.

应该说明的是这种高次多项式拟合不上机操作是非常难实现的,难就难在法方程组大多是病态的,难以控制解的误差.

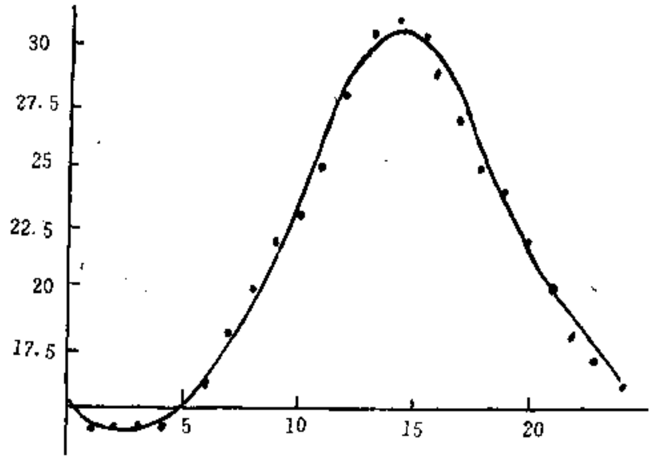


图 3-13 （七次拟合）

例 3 汽车制造厂生产的某种轿车的外形数据如下表 3-15 所示.

表 3-15

x_i	1.16	1.2	1.4	1.6	1.8	2.0	2.2	2.4	2.6	2.8	3.0	3.1	3.2
y_i	0.91	0.95	1.09	1.16	1.19	1.20	1.19	1.17	1.14	1.08	0.99	0.93	0.84

试找出最佳的拟合曲线拟合以上数据

解：利用数学软件包计算进行最佳拟合曲线的选择，由给定的数据的特点，用多项式拟合且是二次以上的多项式来进行数值实验.

```
In[1]:=a={{1.16,0.91},{1.2,0.95},{1.4,1.09},{1.6,1.16},{1.8,1.19},
        {2.0,1.20},{2.2,1.19},{2.4,1.17},{2.6,1.14},{2.8,1.08},
        {3.0,0.99},{3.1,0.93},{3.2,0.84}};
In[2]:=b=ListPlot[a];
In[3]:=c=Fit[a,{1,x,x^2},x]
out[3]=0.089 74+1.023 34x-0.234 787x^2
In[4]:=d=Plot[c,{x,1.1,3.2}];
In[5]:=Show[b,d]
out[5]=-Graphics-
In[6]:=e=Fit[a,{1,x,x^2,x^3},x]
out[6]=-1.261 53+3.117 82x-1.249 7x^2+0.154 586x^3
In[7]:=g=Plot[e,{x,1.1,3.2}];
In[8]:=Show[b,g]
out[8]=-Graphics-
In[9]:=h=Fit[a,{1,x,x^2,x^3,x^4},x]
out[9]=0.085 043 1+0.307 132x+0.846 517x^2-0.509 797x^3+0.075 903 5x^4
In[10]:=k=Plot[h,{x,1.1,3.2}];
```

```
In[11]:=Show[b,k]
out[11]=-Graphics-
In[12]:=w=Fit[a,Table[x^i,{i,0,5},x]]
out[12]=-13.786 7+36.197 3-35.017 2x^2+16.825 6x^3-3.987 35x^4+0.370
        44x^5
In[13]:=v=Plot[w,{x,1.1,3.2}];
In[14]:=Show[b,v]
out[14]=-Graphics-
```

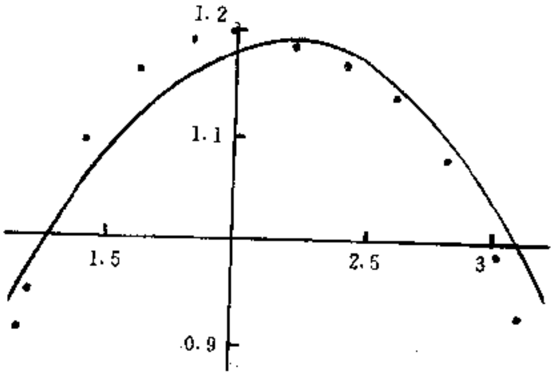


图 3-14 （二次拟合）

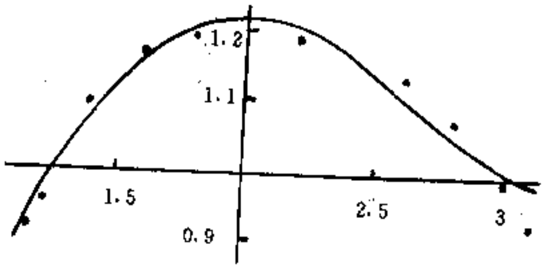


图 3-15 （三次拟合）

从上可知五次拟合为最佳的拟合曲线.

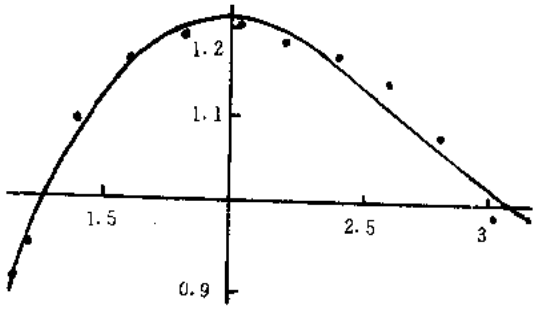


图 3-16 （四次拟合）

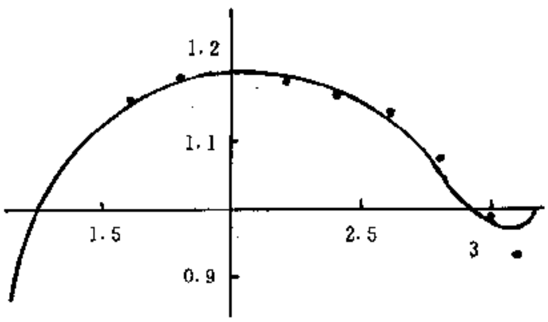


图 3-17 （五次拟合）

二、线性最小二乘拟合

设 $\varphi_0(x),\varphi_1(x),\cdots,\varphi_n(x)$为 $n+1$ 个线性无关的函数，对给定的数据$(x_i,y_i)(i=0,1,\cdots,m)$，求 $P(x)=a_0\varphi_0(x)+a_1\varphi_1(x)+\cdots+a_n\varphi_n(x)\quad(n<m)$，

使
$$\theta(a_0,\cdots,a_n)=\sum_{i=0}^{m}[p(x_i)-y_i]^2$$
最小，这就是线性最小二乘拟合问题.

显然多项式拟合是线性最小二乘拟合的特例，因此同多项式拟合类似，由多元函数极

值的必要条件，$a_0,\cdots,a_n$ 满足：

$$\frac{\partial\theta}{\partial a_0}=0,\quad \frac{\partial\theta}{\partial a_1}=0,\cdots,\frac{\partial\theta}{\partial a_n}=0,$$

于是得关于 $a_0,\cdots,a_n$ 的线性方程组

$$G^TGa=G^TY \tag{3.33}$$

其中 $a=(a_0,a_1,\cdots,a_n)^T$，　$Y=(y_0,y_1,\cdots,y_n)^T$

$$G=\begin{bmatrix}\varphi_0(x_0) & \varphi_1(x_0) & \cdots & \varphi_n(x_0)\\ \varphi_0(x_1) & \varphi_1(x_1) & \cdots & \varphi_n(x_1)\\ \varphi_0(x_m) & \varphi_1(x_m) & \cdots & \varphi_n(x_m)\end{bmatrix}.$$

式(3.33)称为线性最小二乘问题的法方程组，由于 $\varphi_0,\varphi_1,\cdots,\varphi_n$ 线性无关，故行列式 $|G^TG|\neq 0$，方程(3.33)存在唯一的解.

显然多项式拟合是线性拟合的一个特例，即 $\varphi_0(x)=1,\varphi_1(x)=x,\cdots,\varphi_n(x)=x^n$.

例 3　一个男孩在 11 岁长到 21 岁过程中身高的变化如表 3-16，试找一个最佳的拟合曲线来表示这个男孩的成长过程.

表 3-16

t_i(年龄)(从 11 岁起)	0	0.8	1.4	2.0	2.4	3.2	4.0	4.8	5.4
增长高度 $h_i(cm)$	0	0.74	2.25	5.25	8.25	15.00	21.38	26.25	28.88
t_i	6.0	7.0	8.0	10.0					
h_i	30.60	32.25	33	35					

解：所给数据标在坐标纸上，从图 3-18 中看到取正弦级数为拟合曲线较为合适. 把区间(0,10)向右延伸一倍，使得拟合曲线在(0, 20)上关于 $x=10$ 直线对称，然后对拟合

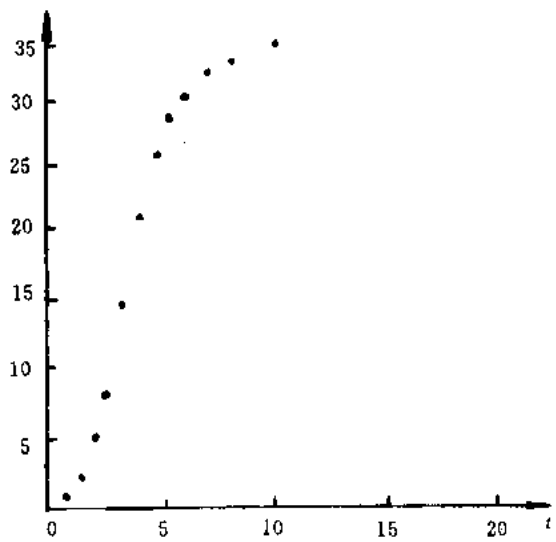


图 3-18

曲线作周期为 40 的奇延拓. 取正弦级数的前三项作为拟合函数有

$$P(t)=b_1\sin\frac{\pi t}{20}+b_3\sin\frac{3\pi t}{20}+b_5\sin\frac{5\pi t}{20}$$

记 $\varphi_0(t)=\sin\frac{\pi t}{20},\varphi_1(t)=\sin\frac{3\pi t}{20},\varphi_2(t)=\sin\frac{5\pi t}{20}$

则此线性拟合问题的法方程组为

$$\begin{bmatrix} \sum_{i=0}^{12}\varphi_0^2(t_i) & \sum_{i=0}^{12}\varphi_0(t_i)\varphi_1(t_i) & \sum_{i=0}^{12}\varphi_0(t_i)\varphi_2(t_i) \\ \sum_{i=0}^{12}\varphi_0(t_i)\varphi_1(t_i) & \sum_{i=0}^{12}\varphi_1^2(t_i) & \sum_{i=0}^{12}\varphi_1(t_i)\varphi_2(t_i) \\ \sum_{i=0}^{12}\varphi_0(t_i)\varphi_2(t_i) & \sum_{i=0}^{12}\varphi_1(t_i)\varphi_2(t_i) & \sum_{i=0}^{12}\varphi_2^2(t_i) \end{bmatrix}\begin{bmatrix} b_1 \\ b_3 \\ b_5 \end{bmatrix}=\begin{bmatrix} \sum_{i=0}^{12}\varphi_0(t_i)h_i \\ \sum_{i=0}^{12}\varphi_1(t_i)h_i \\ \sum_{i=0}^{12}\varphi_2(t_i)h_i \end{bmatrix}$$

解之得：$b_1=35.94$， $b_3=-1.206\,7$， $b_5=-3.507$

拟合函数为：

$$P(t)=35.94\sin\frac{\pi t}{20}-1.206\,7\sin\frac{3\pi t}{20}-3.5077\sin\frac{5\pi t}{20}.$$

从图 3-19 中可以看到拟合曲线较好地拟合了所给的数据.

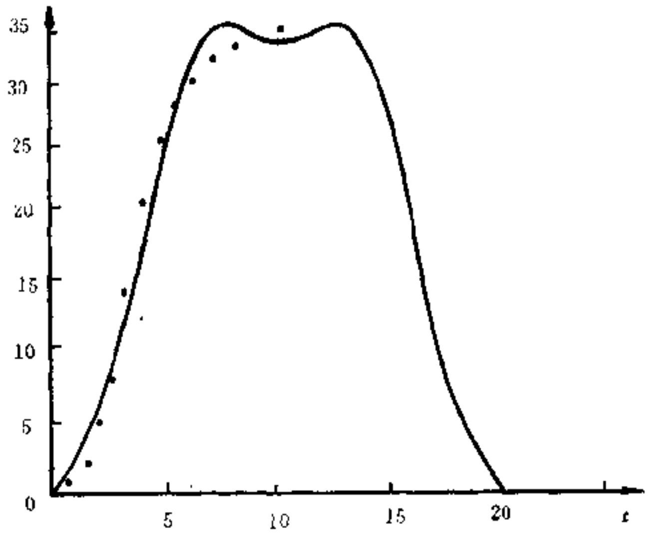


图 3-19

三、非线性拟合

当拟合函数 $P(x)=P(x;a_0,a_1,\cdots,a_n)$，关于参数 $a_0,a_1,\cdots,a_n$ 是非线性时，称为非线性拟合.

对非线性拟合问题的解决方法，最常用的方法是化为线性最小二乘问题. 其它方法限于篇幅，不在此介绍.

例 4 在某化学反应里，根据实验所得生成物的浓度与时间关系如表 3-17，求浓度 y

与时间 t 的拟合曲线 $y=F(t)$.

表 3-17

t_i	1	2	3	4	5	6	7	8
$y_i(\times10^{-3})$	4.00	6.40	8.00	8.80	9.22	9.50	9.70	9.86
t_i	9	10	11	12	13	14	15	16
$y_i(\times10^{-3})$	10.00	10.20	10.32	10.42	10.50	10.55	10.58	10.60

解：将数据标在坐标纸上，我们看到开始时浓度增加较快，后来逐渐减弱，到一定时间就基本稳定在一个数值上．即当 $t\to\infty$ 时，y 趋于某个定数，故有一水平渐近线．$t=0$ 时，反应未开始，生成物的浓度为零，根据这些特点，可设想 $y=F(t)$ 是双曲线或指数型曲线.

(1)取双曲线拟合函数：

$$\frac{1}{y}=a+\frac{b}{t},$$

为了确定 a、b，令

$$\bar{y}=\frac{1}{y},x=\frac{1}{t}$$

于是可用 x 的线性函数 $\bar{y}=a+bx$ 来拟合数据 $(x_i,\bar{y}_i)(i=1,2,\cdots,16)$，数据 $(x_i,\bar{y}_i)$ 由 (t_i,y_i) 变换计算出来，见表 3-18.

表 3-18

x_i	1	1/2	1/3	1/4	1/5	1/6	1/7	1/8
$\bar{y}_i(\times10^3)$	0.250 0	0.156 25	0.125 00	0.113 64	0.108 46	0.105 26	0.103 09	0.101 42
x_i	1/9	1/10	1/11	1/12	1/13	1/14	1/15	1/16
$y_i(\times10^3)$	0.100 00	0.098 04	0.096 90	0.095 97	0.095 24	0.094 79	0.094 52	0.094 34

利用 Mathematica 软件包线性拟合表 3-18 的数据得

```
ln[1]:=a={{1,250},{1/2,156.25},{1/3,125},{1/4,113.64},{1/5,108.46},
     {1/6,105.26}, {1/7, 103. 09}, {1/8, 101. 42}, {1/9, 100},
     {1/10,98.04}, {1/11,96.9}, {1/12,95.97}, { 1/13, 95. 24 },
     {1/14,94.79},{1/15,94.52},{1/16,94.3434}};
ln[2]:=f=Fit[a,{1,x},x]
out[2]=79.818+163.283x
```

即

$$\bar{y}=79.818+163.283x$$

从而得：

$$y=F_1(t)=\frac{t}{79.818t+163.283}$$

(2)取拟合函数为指数型

$$y = ae^{b/t} \quad (b < 0)$$

上式两边取自然对数

$$\ln y = \ln a + b/t$$

令 $\hat{y}=\ln y, \hat{t}=\frac{1}{t}, \hat{a}=\ln a$，则拟合函数化为

$$\hat{y} = \hat{a} + b\hat{t}$$

同上述过程类似，由 (t_i, y_i) 相应算出 $(\hat{t}_i, \hat{y}_i)$，然后利用线性拟合，解得

$$\hat{a} = -4.480\,72, b = -1.057\,669$$

从而得 $a=e^{\hat{a}}=1.132\,53\times10^{-2}$

拟合函数为

$$y = F_2(t) = 1.132\,43 \times 10^{-2}e^{-1.056\,69/t}.$$

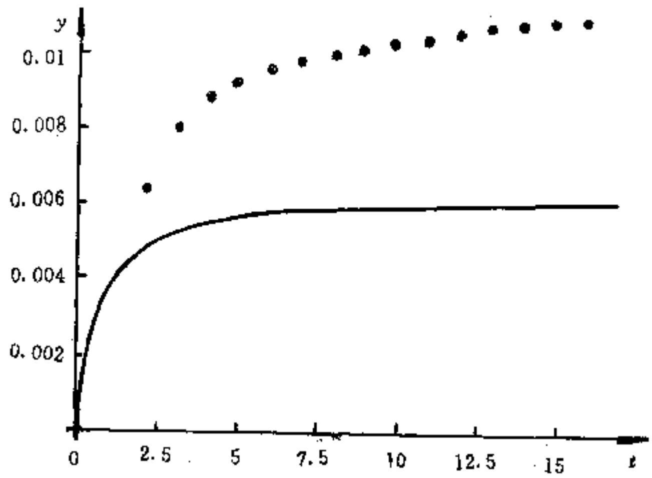


图 3-20

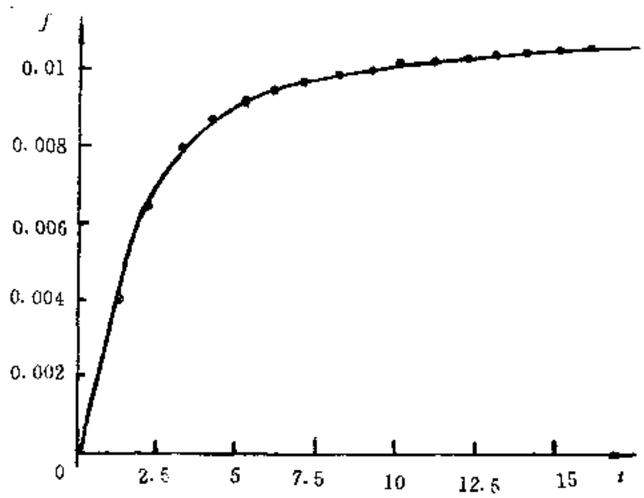


图 3-21

为了比较两个数学模型的好坏，分别作数据点与拟合曲线的叠加图，图 3-20 是双曲线拟合，图 3-21 是指数型拟合.

显然 $F_2(t)$拟合得更好.

可以求得各自的均方误差为

$$\sqrt{\sum_{i=0}^{15}[y_i - F_1(t_i)]^2} - 1.25 \times 10^{-3}, \sqrt{\sum_{i=0}^{15}[y_i - F_2(t_i)]^2} = 0.34 \times 10^{-3},$$

四、交通事故调查

发生严重事故之后,警方总希望能对肇事车辆的速度有个估计,尽管在场的司机能够提供这一数据,但警方还是希望能有一个建立在自已观测之上的有说服力的证据.在合适的条件下,路面上会留下刹车的痕迹,这些痕迹的外观可以告诉我们车轮是怎样打滑的,黑色的,造成一线的痕迹表明刹车时车轮已被闸住不能转动.车子因惯性向前沿直线冲出,形如人字呢纹路的痕迹则表明车轮仍在转动,但车身滑向路边.下面给出两个交通事故的调查问题:

问题 I

一辆汽车驶向一个刚刚变成红色的交通信号灯,只听吱的一声,汽车停住,在路上留下了一条滑痕.一名闻声赶来的警察看了看地下的刹车痕迹,指责司机刚才超速行驶.

请提供一个使警察能够计算该车制动前行驶速度的办法.

1.问题的模型

汽车停下来是由于轮胎与路面之间的摩擦力,假设在整个刹车过程中这个摩擦力是个常量,用 μ 表示摩擦系数,由牛顿第二定律知刹车过程中的减速度可以表示为:

$$a = \mu g (g\text{ 是重力加速度})$$

由于减速度是常量,从而

$$v^2 = v_0^2 - 2as$$

由于汽车最后是完全停了下来,故 $v=0$,于是

$$v_0^2 = 2as$$

即

$$v_0^2 = ks$$

其中 $k=2a=2\mu g$,v_0 为初速度(以英里/小时为单位),s 为制动距离(以英尺为单位).

要确定汽车的速度,就必须知道 a 或 μ.警方的有关部门针对不同的天气条件和不同的路面情况编制了摩擦系数的表,见表 3-19.

不过更可取的方法是在事故现场实地测出 a 或 μ 的值.测量的办法之一是在车上装一台指示加速度的仪表,这样,在急刹车时就可以从表上直接读出加速度来,不过这方法并不常用.比较通行的作法是使一个汽车在已知的速度下刹车,测出在地面滑动的距离,并由此计算出 a 值,这种作法有一个优点,那就是对具有任何坡度的倾斜路都能应用,警察一般作两次刹车试验,然后根据较长的那条刹车痕迹来计算减速度.一般来说,确定痕迹的地点比较困难,对事故现场要求十分仔细,所以这种以警察的观测得出的最后结论会有 5%的误差.

表 3-19　橡胶轮胎与路面间阻力系数的区间

路面情况		干				湿			
		30 英里/小时以下		30 英里/小时以上		30 英里/小时以下		30 英里/小时以上	
		区	间	区	间	区	间	区	间
混凝土路	新、糙路面	0.80	1.00	0.70	0.85	0.50	0.80	0.40	0.75
	一般路面	0.60	0.80	0.60	0.75	0.45	0.70	0.45	0.65
	旧路面	0.55	0.75	0.50	0.65	0.45	0.65	0.45	0.60
柏油路面	新、糙路面	0.80	1.00	0.65	0.70	0.50	0.80	0.45	0.75
	一般路面	0.60	0.80	0.55	0.70	0.45	0.70	0.40	0.65
	旧路面	0.55	0.75	0.45	0.65	0.45	0.65	0.40	0.60
	超细柏油路	0.50	0.60	0.35	0.60	0.30	0.60	0.25	0.55
砖路	新路面	0.75	0.95	0.60	0.85	0.50	0.75	0.40	0.70
	旧路面	0.60	0.80	0.55	0.75	0.40	0.70	0.40	0.60
石板路	新路面	0.75	1.00	0.70	0.90	0.60	0.90	0.60	0.85
	旧路面	0.50	0.70	0.45	0.65	0.30	0.50	0.25	0.50
碎石路	灌浆、浸油	0.55	0.85	0.50	0.80	0.40	0.80	0.40	0.60
	松软	0.40	0.70	0.40	0.70	0.45	0.75	0.45	0.75
煤渣路	灌浆	0.50	0.70	0.50	0.70	0.65	0.75	0.65	0.75
岩石路	碾碎	0.55	0.75	0.55	0.75	0.55	0.75	0.55	0.75
冰路	平滑	0.10	0.25	0.07	0.20	0.05	0.10	0.05	0.10
雪路	化浆	0.30	0.55	0.35	0.55	0.30	0.60	0.30	0.60
	松软	0.10	0.25	0.10	0.20	0.30	0.60	0.30	0.60

问题 Ⅱ

一辆汽车在拐弯时急刹车，结果冲进路边的沟里，警察闻讯赶到现场，对汽车留在路上的刹车痕迹进行了细致的测量，利用所测到数据画出了事故现场的平面图，如图 3-22 所示.

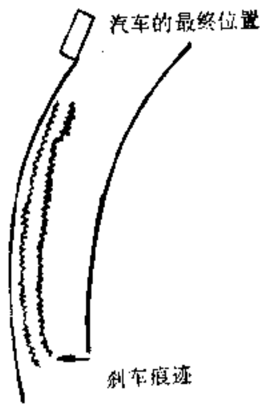


图 3-22

司机说：当车进入弯道时刹车失灵，且进入弯道时的车速为 40 英里/小时（该路的速度上限）. 通过验车证实该车的制动器在事故发生时的确失灵. 但司机所说的车速是否真实呢？

请给出上个使警察可以核对速度的计算办法.

通常，作一条基准线测量刹车痕迹. 距离 x 沿基准线测得，距离 y 与 x 垂直(见图 3-23)表 3-20 给

出了外测刹车痕迹的有关值(单位:米).

表 3-20

x	0	3	6	9	12	15	16.64	18	21	24	27	30	33.27
y	0	1.19	2.15	2.82	3.28	3.53	3.55	3.54	3.31	2.89	2.22	1.29	0

警察还测量了路的坡度,发现这段路是平的.通过痕迹图可得这样的一个信息:该车并没有偏离它行驶的转弯曲线,也就是说车头一直指向切线方向.

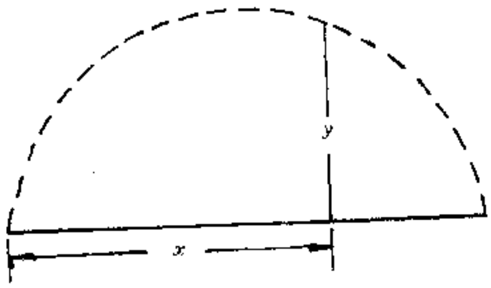


图 3-23

问题 Ⅱ 的模型:

汽车沿弯路行驶,车轮转着打滑,车滑向路边.假定摩擦力作用在汽车速度的法线上.设汽车速度的 v 是个常量,并设汽车重心沿一个半径为 r 的圆运动.在这种情形下,摩擦力提供了向心力.用 μ 表示摩擦系数,则

$$\mu g = \frac{v^2}{r}$$

从而

$$v = \sqrt{\mu g r}$$

μ 的测量方法在问题 I 中已经讨论过了,μ 测试后就能算出减速度 a:

$$a = \mu g$$

在这种情形下,将结果写成

$$v = \sqrt{ar}$$

更方便些.

从上面的讨论可知,速度 v 的求法最后归结为求圆的半径 r.为此可用表 3-14 的数据拟合一个圆,设该圆的圆心坐标为 (x_p, y_p),要求

$$\sum_{i=0}^{12} \left| (x_i - x_p)^2 + (y_i - y_p)^2 - r^2 \right|$$

最小.

下面给出用最小二乘法求解的方法

设近似圆弧为

$$(x - x_p)^2 + (y - y_p)^2 = r^2,$$

展开得

$$x^2 + y^2 - 2x_p x - 2y_p y + x_p^2 + y_p^2 - r^2 = 0,$$

令

$$\begin{cases} 2x_p = -a_1, \\ 2y_p = -a_2, \\ x_p^2 + y_p^2 - R^2 = a_0. \end{cases}$$

解得

$$\begin{cases} x_p = -\dfrac{a_1}{2} \\ y_p = -\dfrac{a_2}{2} \\ r = \dfrac{1}{2}\sqrt{a_1^2 + a_2^2 - 4a_0} \end{cases}$$

将上两式代入圆弧中得

$$a_0 + a_1 x + a_2 y + x^2 + y^2 = 0$$

问题仍归结为求 a_0, a_1, a_2. 将表 3-14 中的 13 对数据 $(x_2, y_i)(i=0,1,\cdots,12)$ 代入上式，则其分别得出的偏差为

$$\triangle_i = a_0 + a_1 x_i + a_2 y_i + x_i^2 + y_i^2 \quad (i = 0,1,\cdots,12)$$

其误差平方和为

$$\varepsilon = \sum_{i=0}^{12} \Delta_i^2 = \sum_{i=0}^{12} (a_0 + a_1 x_i + a_2 y_i + x_i^2 + y_i^2)^2$$

由使 ε 最小的极值条件得：

$$\begin{cases} \dfrac{\partial \varepsilon}{\partial a_0} = \sum\limits_{i=0}^{12} 2(a_0 + a_1 x_i + a_2 y_i + x_i^2 + y_i^2) = 0 \\ \dfrac{\partial \varepsilon}{\partial a_1} = \sum\limits_{i=0}^{12} 2x_i(a_0 + a_1 x_i + a_2 y_i + x_i^2 + y_i^2) = 0 \\ \dfrac{\partial \varepsilon}{\partial a_2} = \sum\limits_{i=0}^{12} 2y_i(a_0 + a_1 x_i + a_2 y_i + x_i^2 + y_i^2) = 0 \end{cases}$$

整理后写成矩阵形式为

$$\begin{bmatrix} 13 & \sum\limits_{i=0}^{12} x_i & \sum\limits_{i=0}^{12} y_i \\ \sum\limits_{i=0}^{12} x_i & \sum\limits_{i=0}^{12} x_i^2 & \sum\limits_{i=0}^{12} x_i y_i \\ \sum\limits_{i=0}^{12} y_i & \sum\limits_{i=0}^{12} x_i y_i & \sum\limits_{i=0}^{12} y_i^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} -\left(\sum\limits_{i=0}^{12} x_i^2 + \sum\limits_{i=0}^{12} y_i^2\right) \\ -\left(\sum\limits_{i=0}^{12} x_i^3 + \sum\limits_{i=0}^{12} x_i y_i^2\right) \\ -\left(\sum\limits_{i=0}^{n} x_i^2 y_i + \sum\limits_{i=0}^{n} y_i^3\right) \end{bmatrix},$$

代入表 3-14 的数据得如下的三阶线性方程组

$$\begin{bmatrix} 13 & 214.91 & 29.77 \\ 214.91 & 4\,848.782\,5 & 494.462 \\ 29.77 & 494.462 & 88.245\,1 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} -4\,937.027\,6 \\ -124\,577.131\,6 \\ -9\,892.662\,647 \end{bmatrix}.$$

利用 Mathematica 软件解此方程组

```
In [1]: = A = {{13, 214, 91, 29. 77}, {214  91, 4  848. 782  5, 494. 462},
        {29.77,494.462,88.245 1}};
```

```
In[2]:=b={-4 937.027 6,-124 577.131 6,-9 892.662 647};
In[3]:=LinearSolve[A,b]
out[5]={0.397 327,-33.281 5,74.247 4}
```

即

$$a = 0.397\ 327, a_1 = -33.281\ 5, a_2 = 74.247\ 4$$

从而得圆坐标为(16.64,-37.12).

由半径 $r=\frac{1}{2}\sqrt{a_1^2+a_2^2+4a_0}$ 得

$$r = 40.677\ 8$$

有了半径的资料及路面的摩擦系数,我们就可以计算出汽车的速度 v,从而确定司机关于车速的说法是否真实.

此模型也可以推广到坡路的情况,弯道的路面是向内倾斜的(有利于转弯)情况等.

五、估计水塔的水流量

1. 问题的提出

美国某州的各用水管理机构要求各社区提供每小时多少加仑的用水率以及每天所用的总水量.许多社区没有测量流入或流出当地水塔的水量的装置,他们只能代之以每小时测量水塔中的水位,其精度不超过0.5%,更重要的是,当水塔中的水位下降到最低水位 L 时,水泵就启动向水塔输水直到最高水位 H,但也不能测量水泵的供水量,因此,当水泵正在输水时,不容易建立水塔中水位和水泵工作时用水量之间的关系.水泵每天输水一次或二次,每次约二小时.

表 3-21 某小镇某天水塔水位

时间(秒)	水位(0.01 英尺)	时间(秒)	水位(0.01 英尺)
0	3 175	46 636	3 350
3 316	3 110	49 953	3 260
6 635	3 054	53 936	3 167
10 619	2 994	57 254	3 087
13 937	2 947	60 574	3 012
17 921	2 892	64 554	2 927
21 240	2 850	68 535	2 842
25 223	2 795	71 854	2 767
28 543	2 752	75 021	2 697
32 284	2 692	79 254	水泵开动
35 932	水泵启动	82 649	水泵开动
39 332	水泵启动	85 968	3 475
39 435	3 550	86 968	3 397
43 318	3 445	92 370	3 340

试估计任何时刻(包括水泵正在输水的时间内)从水塔流出的水流量 $f(t)$,并估计出一天的总用水量.下表给出了某个小镇一天中真实的数据.

表中给出了从第一次测量开始的以秒为单位的时刻，以及该时刻的高度单位为百分之一英尺的的水位测量值．例如，3 316 妙后，水塔中水位达到 31.10 英尺，水塔是一个高为 40 英尺，直径为 57 英尺的正圆柱，通常当水塔水位降至约为 27.00 英尺时，水泵开始工作，当水位升到 35.50 英尺时水泵停止工作．

2．模型的假设

(1)水塔中的流出的水流量只受社区的日常生活需要的影响，即表 3-21 中所给的数据不反映非正常需要，如水塔泄漏，主水管破裂，火灾救火，自然灾害等．

(2)由 Forrilli 定律知，从水塔流出的最大流速正比于水位高度的平方根，从所给的数据可知最大高度为 35.5 英尺，最小高度为 27 英尺，所以对于这两种高度，最大水流速度的比约为$\sqrt{35.5/27}\approx 1.15$．这一点表明我们可以假设水塔中的水位不会太多地影响水流速度，同样地，我们假设天气条件、温度改变不直接影响水流速度．

(3)水泵的输水速率是个常数．没有证据证明水泵输水速率是随时间变化的，或是前面输入水量的函数．我们假设水泵在水塔水位大约在 27 英尺时开始起动，在大约 35.5 英尺时关闭．另外，水泵不会中断或需要维修．

(4)从水泵中流出的水流的最大流速要比水泵注水的速度小．这是因为社区一直需要用水，不允许水塔的水用光．

(5)水的消耗每天大致差不多．这是因为水的消耗依赖于人们的日常生活，如洗澡、烧饭、洗衣等．

(6)从水塔中流出的水流流速的改变率可以用一条光滑曲线来逼近．这是因为虽然就个别用户而言，可能用水量有较大的变化，但由于个人的用水量与整个社区需求量相比是非常小的，从统计意义上来讲，不太可能同时整个社区的用水量增长或减少．

为了方便起见，我们引入以下的记号：

V 表示水量(加仑)，V_t 表示 t 时刻水量(加仑)，

t 时间(小时)，$f(t)$是由模型所估出的 t 时刻水塔流出的水量(加仑/小时)，

p 表示水泵的注水速率(加仑/小时)

3．问题的分析

水泵两次注水，但水泵的真正开启和停机的时刻我们并不知道．因为水泵启动前，测得水位两次均为26.97 英尺，可以认为水泵开启时间为测量后不久，而水泵停止工作后紧接着测量所得的水位不同于水泵刚刚停止工作的水位．所以两次测量之间水泵一定已经停止工作．图 3-24 是所给水量(加仑)数据关于时间(小时)的函数，得到这个函数关系应清楚：

1 英尺＝3.048 分米，

1 加仑(美制)＝3.785 公升．

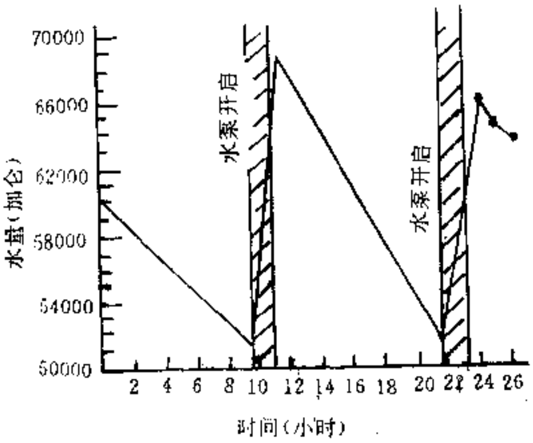


图 3-24

从而得到:1 立方英尺=7.484 加仑(美制).

下面估计水泵注水的起始和终止时刻.

第一次水泵开动时间在 8.968 小时(即 32 284 秒)后不久,因为这时的水量大约为 $V_i = \pi\times\left(\frac{57}{2}\right)^2\times 26.97\times 7.484\approx 514\ 800$ 加仑.水泵停止工作的时间为 10.954 小时(39 435 秒).因为 10.954 小时测得的水位为 35.50 英尺.

对于第二次水泵开启时间,可以认为 20.839 小时(即 75021 秒),因为这时的水量大约也在 514,800 加仑.我们选终止时刻为 22.858 小时,因为水泵停止工作后的第一次观测点为(23.880 小时),比第一次水泵停时的水位少 14.300 加仑.

表 3-22 显示了水流与时间(相邻时间段的中点)之间的关系,图 3-25 显示的是表 3-22 的数据.下面用函数 $f(t)$来拟合这些数据,之后再对 $f(t)$积分,用以计算每天的用水量.

表 3-22　观测时间中点与该时间区间平均水流关系

时间中点(小时)	平均流速($\times10^3$ 加仑/小时)	时间中点(小时)	平均流速($\times10^3$ 加仑/小时)
0.460 4	14.0	13.42	19.0
1.382	12.0	4.43	16.0
2.396	10.0	1.44	16.0
3.411	9.6	16.37	16.0
4.425	9.6	17.38	14.0
5.439	8.9	18.49	14.0
6.453	9.6	19.5	16.0
7.468	8.9	20.40	15.0
8.448	10.0	21.43	*no data*
9.474	*no data*	22.49	*no data*
10.45	*no data*	22.49	*no data*
11.49	18.6	24.43	14.0
12.49	20.0	25.45	12.0

4.模型建立和检验

我们用 7 次多项式拟合图 3-25 中的数据,用数学软件计算得

$$f(t) = 12\ 499.8 + 2\ 773.89t + 2\ 466.42t^2 - 556.628t^3 - 53.33t^4 + 2.498\ 16t^5 - 0.056\ 149\ 4t^6 + 0.000\ 478\ 418t^7$$

t 的定义如下

$$t = \begin{cases} 24 + (T - T_0) + 0.460\ 6, & 对\ T < T_0, \\ (T - T_0) + 0.460\ 6, & 对\ T \geqslant T_0, \end{cases}$$

其中 T_0 表示用来作为拟合的初始数据点的时刻(小时),T 表示一天中的时间(在 24 小时之内).

图 3-26 是表 3-22 的数据点与拟合曲线 $f(t)$ 的叠加图，从中看到拟合的情况. 刚开始接触该问题，曾想利用超过 7 次的多项式拟合数据，然而高次多项式的拟合会出现严重的波动，之所以不选用低于 7 次的多项式是考虑到均方误差.

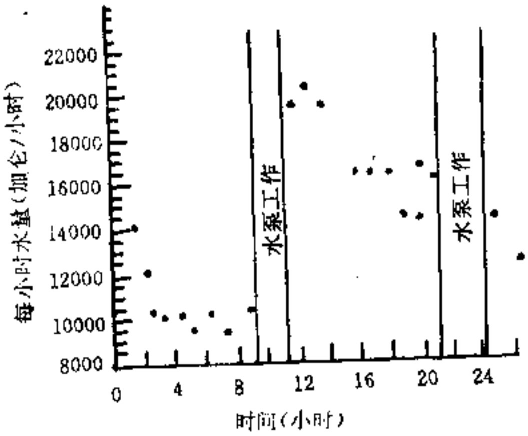


图 3-25

我们假设水泵的注水速度是常数，因为在水泵注水时间内水量的平均改变率 $\Delta V/\Delta t$ 应该近似地等于注水速度 p 减去水塔在此期间内的平均流出速度，所以 p 可表达为

$$p = \frac{\Delta V}{\Delta t} + \frac{\int f(t)\mathrm{d}t}{\Delta t}$$

这里，$\int f(t)\mathrm{d}t$ 表示在 Δt 区间内积分.

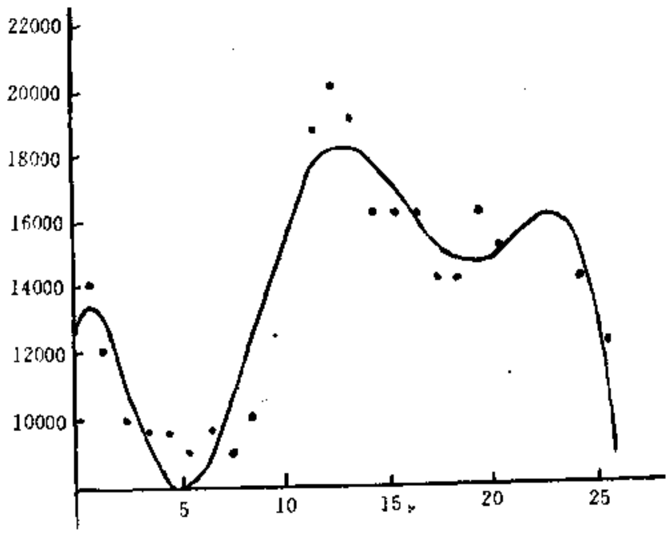


图 3-26

如果实际的模型能准确地模拟数据，那么在不同的注水期间，水泵的注水速度应保持为常数. 下面利用两个不同的起始和终止的例子来检验这一点.

$$\Delta V_1 = 667\ 600 - 514\ 800 = 162\ 800 \text{ 加仑},$$

$$\Delta t_1 = 10.954 - 8.968 = 1.986 \text{ 小时},$$

$$\frac{\Delta V_1}{\Delta t_1}=81\ 973 \text{ 加仑 / 小时；}$$

$$\Delta V_2 = 667\ 600 - 514\ 800 = 162\ 800 \text{ 加仑，}$$

$$\Delta t_2 = 22.858 - 20.839 = 2.019 \text{ 小时，}$$

$$\frac{\Delta V_2}{\Delta t_2}=80\ 634 \text{ 加仑 / 小时；}$$

在 Δt_1 时间内 $f(t)$的均值为：

$$\frac{\int_{8.968}^{10.954} f(t)\mathrm{d}t}{\Delta t_1} = 15\ 245 \text{ 加仑 / 小时，}$$

所以，水泵第一次注水速率为

$$p_1 = 81\ 973 + 15\ 245 = 97\ 218 \text{ 加仑 / 小时.}$$

在 Δt_2 时间内 $f(t)$的均值为

$$\frac{\int_{20.839}^{22.858} f(t)\mathrm{d}t}{\Delta t_2} = 15\ 359 \text{ 加仑 / 小时，}$$

所以，水泵第二次注水速率为

$$p_2 = 80\ 634 + 15\ 359 = 95\ 993 \text{ 加仑 / 小时.}$$

计算 p_i 的相对误差仅为 1.6%，这证明了模型的合理性.

5. 结论

因为在 0.460 与 24.460 点 $f(t)$的值近似相等(13 506 加仑/小时，14 162 加仑/小时，相对误差为 3%)，所以 f 可以看作是一个连续的周期函数.

一天的总用量可以用 $f(t)$曲线在 24 小时内的积分，即一个周期所用的面积来表示，计算得

$$\int_{0.460\ 6}^{24+0.460\ 6} f(t)\mathrm{d}t = 334\ 943 \text{ 加仑.}$$

由 $f(t)$我们确定水塔流出的水流速度为

$$\frac{1}{24}\int_{0.460\ 6}^{24+0.460\ 6} f(t)\mathrm{d}t \approx 13\ 956 \text{ 加仑 / 小时.}$$

另外，水泵的平均注水速度为

$$(97\ 218 + 95\ 993)/2 \approx 96\ 606 \text{ 加仑 / 小时.}$$

6. 模型的优缺点

(1)优点

①这个模型的灵活性很好，并可以用到那些拥有竖直圆柱型水塔的小城镇. 另外，模型中的输入数据可以是在任何近似均匀间隔的水位高度.

②模型中的数学概念简单，并且容易理解.

(2)缺点

我们的模型受到水塔的几何形状的限制，只允许用垂直水塔.

习题三

一、给定函数表

x	0.10	0.15	0.25	0.30
e^{-x}	0.904 8	0.860 7	0.778 8	0.740 8

用线性插值和抛物插值计算 e^{-x}在 $x=0.23$ 处的近似值，并估计截断误差.

二、给出 $f(x)$的函数表

x	0.40	0.55	0.65	0.80	0.90	1.05
$f(x)$	0.410 75	0.578 15	0.696 75	0.888 11	1.026 52	1.253 82

求 4 次牛顿插值多项式，并由此计算 $f(0.596)$的近似值.

三、已经测得某地大气压强随高度变化的一组数据.

高度(m)	0	300	600	1 000	1 500	2 000
压强(kgf/m²)	0.968 9	0.932 2	0.896 9	0.851 9	0.798 9	0.749 1

试求出每隔 100m 的压强值，并与理论计算公式

$$P = 1.033\,2e^{-\frac{x+500}{7\,756}}$$

的计算结果作比较

四、在微电机设计计算中需要查磁化曲线表，通常给出的表是磁密 B 每间隔 100 高斯磁路每厘米长所需安匝数 at 的值，试解决 B 从 400 至 1100 区间的查表问题. 已知数据如下：

k	0	1	2	3	4	5	6	7
B_k	4 000	4 500	5 000	5 500	6 000	6 500	7 000	7 500
$at_k=f(B_k)$	1.38	1.48	1.58	1.69	1.81	1.94	2.10	2.28
k	8	9	10	11	12	13	14	
B_k	8 000	8 500	9 000	9 500	10 000	10 500	11 000	
$at_k=f(B_k)$	2.50	2.76	3.06	3.41	3.83	4.33	4.93	

五、已知 $y=f(x)$的函数表如下

x	−1	0	1	3
y	−1	1	3	31
y'	4			28

求其三次样条插值函数 $s(x)$，并用 $s(x)$ 求 $f(-0.5)$ 和 $f(2)$ 的近似值.

六、从人口普查统计，已知某国新生儿的母亲的年龄累计分布为 $N=N(t)$，其中 t 为母亲年龄，N 为新生儿的母亲年龄低于或等于 t 的新生儿数目. 下表给出了间距为 5 的数据：

t	15	20	25	30	35	40	45	50
N	0	7 442	26 703	41 636	49 785	50 209	50 226	50 230

取边界条件为 $N'(15)=N'(50)=0$，利用三次样条插值求出 15 至 50 岁之间每一年龄(即 $x=15,16,\cdots,50$)所对应的新生儿累计数目 N，并绘出插值函数的图形. (数值实验)

七、用最小二乘法求一个形如 $y=a+bx^2$ 的经验公式，使它与下列数据相拟合. 并计算均方误差

x_i	19	25	31	28	44
y_i	19.0	32.2	49.0	73.0	97.8

八、观测物体的直线运动，得出以下数据：

时间 t(秒)	0	0.9	1.9	3.0	3.9	5.0
距离 s(米)	0	10	30	50	80	110

求运动方程.

九、在某化学反应里，根据实验所得分解物的浓度与时间关系如下：

时间 t(分)	0	5	10	15	20	25	30	35	40	45	50	55
浓度 $y(\times 10^{-4})$	0	1.27	2.16	2.86	3.44	3.87	4.15	4.37	4.51	4.58	4.62	4.64

用最小二乘拟合求 $y=F(t)$.

十、通过分析一个物体在大气中的运动，可以得到海平面以上 z 高度处的大气密度 ρ，试从给出的密度——高度数据中找出经验公式并分析.

高度(z)	0	304.8	609.6	1 219.2	1 828.8	3 048	4 572
密度(ρ)	1.224	1.189	1.154	1.090	1.023	0.904	0.771
高度(z)	6 096	9 144	12 192	15 240	18 288		
密度(ρ)	0.653	0.459	0.302	0.187	0.115		

第四章　数学规划模型

数学规划模型建立在大量的实际问题的基础上，尤其在管理实践中，如何合理安排一定量的人力、物力、财力，使产值最大，效益最高或产品数量最多？怎样才能使库存量即满足需要，又使贮存费用最少？这些问题都要求我们利用数学这一工具，给出一个较为满意的答复．这里我们仅用一些实际例子，采用较简单的方式来说明数学规划模型的建立及求解方法，而不去面面俱到地研究各种算法的理论问题．

第一节　线性规划模型

一、生产计划安排

例 1.1　一个煤矿生产褐煤、无烟煤两种煤．所生产的煤全部都能卖掉．卖一吨褐煤和无烟煤分别获利 4 百元和 3 百元，采煤机、选煤机、洗煤机一天工作时间分别不超过 12、10、8 小时．生产一吨褐煤需用采煤机、选煤机、洗煤机各工作 3、3 和 4 小时，生产一吨无烟则需用采煤机、选煤机、洗煤机各工作 4、3 和 2 小时，问每天各生产两种煤多少吨，才能使煤矿获利最大？

分析：为了解决上述问题，可定义下列变量．

x_1——每天生产褐煤的吨数；

x_2——每天生产无烟煤的吨数；

Z——每天生产 x_1 吨褐煤和 x_2 吨无烟煤所获利润．

对此问题可列出如下方程：

每天可获利

$$Z = 4x_1 + 3x_2$$

并且必须满足下列条件：

采煤机每天工作时间

$$3x_1 + 4x_2 \leqslant 12$$

选煤机每天工作时间

$$3x_1 + 3x_2 \leqslant 10$$

洗煤机每天工作时间

$$4x_1 + 2x_2 \leqslant 8$$

两种煤的产量为正值，即

$$x_1 \geqslant 0, x_2 \geqslant 0$$

归结为使目标函数 Z 取最大值的问题：

$$\max \quad Z = 4x_1 + 3x_2$$

约束 $3x_1 + 4x_2 \leqslant 12$

$3x_1 + 3x_2 \leqslant 10$

$4x_1 + 2x_2 \leqslant 8$

$x_1 \geqslant 0, x_2 \geqslant 10$

二、混料问题的最优设定

例 1.2 在某种饲料的混合搭配中，各种原料的成分及成本如表(表 5-1)．表中三种原料混合，各成份应满足下列要求：

$$a_1 \geqslant 0.04 \qquad a_2 \geqslant 0.02 \qquad a_3 \geqslant 0.07$$

表 4-1

原料	成份			成本(元/公斤)
	a_1	a_2	a_3	
1	0.06	0.02	0.09	15
2	0.03	0.04	0.05	12
3	0.04	0.01	0.03	8

问上述三种原料各点多少，则既能满足成份要求，又能使成本最低．

下面建立这个问题的数学模型．

设原料 1，2，3 各占 x_1, x_2, x_3 份，则有 $x_1+x_2+x_3=1$

其中各项成份应满足条件：

$a_1=0.06x_1+0.03x_2+0.04x_3\geqslant0.04$

$a_2=0.02x_1+0.04x_2+0.01x_3\geqslant0.02$

$a_3=0.09x_1+0.05x_2+0.03x_3\geqslant0.07$

$x_1\geqslant0, x_2\geqslant0, x_3\geqslant0$

且使目标函数取最小值

$$Z = 15x_1 + 12x_2 + 8x_3$$

将 $x_3=1-x_1-x_2$ 代入上面式子中，我们就得到了满足约束条件的最小值问题．

$$\min \quad Z = 15x_1 + 12x_2 + 8(1 - x_1 - x_2)$$

约束 $2x_1 - x_2 \geqslant 0$

$x_1 + 2x_2 \geqslant 1$

$$3x_2 + x_2 \geqslant 2$$

$$x_1 \geqslant 0, x_2 \geqslant 0, 1 - x_1 - x_2 \geqslant 0$$

三、线性规划的一般性问题

1. 线性规划的标准型

对前面所表示的问题，一般可以化成如下线性规划问题的形式：

$$\max \quad Z = ax_1 + c_2x_2 + \cdots + c_nx_n \tag{4.1}$$

$$约束 \quad a_{11}x_1 + _{12}x_2 + \cdots + a_{1n}x_n \leqslant b_1$$

$$\begin{cases} a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \leqslant b_2 \\ \cdots \quad \cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \leqslant b_m \\ x_i \geqslant 0 \quad (i = 1, 2, \cdots, n) \end{cases} \tag{4.2}$$

上述问题的矩阵表示形式为：

$$\max \quad Z = C^TX$$

$$约束 \quad AX \leqslant B$$

$$X \geqslant 0$$

这里 $C = (c_1, c_2, \cdots, c_n)^T$

$X = (x_1, x_2, \cdots, x_n)^T$

$A = (a_{ij})_{m \times n}$

$B = (b_1, b_2, \cdots, b_m)^T$

$0 = (0)_{1 \times n}$

如果目标函数是极小，则有：

$$\min \quad Z = C^TX$$

$$约束 \quad AX \geqslant B$$

$$X \geqslant 0$$

在解线性规划问题时，我闪经常把线性规划的表达式整理成一种标准型：

$$\max \quad Z = C^TX \tag{4.3}$$

$$约束 \quad AX = B \tag{4.4}$$

$$X \geqslant 0 \tag{4.5}$$

其中　$B \geqslant 0$

这种标准型有以下的特征：

(1)目标函数都是求极大；

(2)除非负条件外，约束都是严格等式；

(3)自变量都是非负的；

(4)每个约束方程右端的常数都是非负的；

2. 非标准型化为标准型的方法

线性规划问题只有两类，一类是求最大值问题，一类是求最小值问题. 因此，对于我们所规定的标准型来说，有下面的问题需要解决，并将非标准型的线性规划问题化为标准型.

(1)将求最小值问题化为求最大值问题.

将目标函数的所有系数改变符号，且约束条件不变. 即 $min \quad Z=C^TX$ 化为 $\max \quad \overline{Z} = -Z = (-C)^T X$，求出 $\overline{Z}$ 后，$Z=-\overline{Z}$ 便是最小值，X 不变.

(2)将不等式约束化为等式约束

对于约束为"≤"时，在方程左端加一个非负变量，使其成为等式，新加变量称为松驰变量.

对于约束为"≥"时，在方程左端减去一个非负变量，使其成为等式，这个减去的变量称为剩余变量.

(3)决策变量无非负限制时，化成标准型的方法.

当决策变量 x_i 无非负限制时，则用两个非负变量的差来代替. 即令 $x_i = x'_i - x''_i$，并要求 $x'_i \geqslant 0, x''_i \geqslant 0$.

3. 线性规划问题的解

设线性规划标准型(式 4.3)～(式 4.5)中，有 m 个约束方程，n 个决策变量.

(1)可行解

满足约束条件(式 4.4)～(式 4.5)的解 $X=(x_1, x_2, \cdots, x_n)^T$ 称为线性规划问题的可行解. 所有可行解的集合称为可行域.

(2)最优解

满足(4.1)的可行解称为线性规划问题的最优解(即使目标函数达到最大值的可行解叫最优解).

(3)基本解

在 n 个变量中，若把 $(n-m)$ 个变量置于零，所得的 m 个线性方程与 m 个未知数的方程组有唯一解(系数矩阵是满秩的)，则此唯一解为基本解.

未置零的变量叫做基变量. 相应于基变量的系数阵称为线性规划的一个基.

已置零的变量称为非基变量.

(4)基本可行解

满足约束条件和非负条件的基本解，称为基本可行解.

(5)最优基本可行解

使目标函数达到最优的基本可分解，称为最优基本可行解.

四、线性规划问题的图解法

对简单的线性规划问题，利用图解的方法求最优解是可行的，可以按以下步骤来进行.

首先，根据约束条件找出可行域. 其次，根据目标函数值不同时，函数有一系列斜率相同的平行等高线. 最后，向右平行移动等高线，使之移动到将要离开可行域，并与之相切的点（一般情况下是可行域的顶点），这一切点即为所求的最优解所对应的点.

图 4-1 是例 1.1 的图解过程.

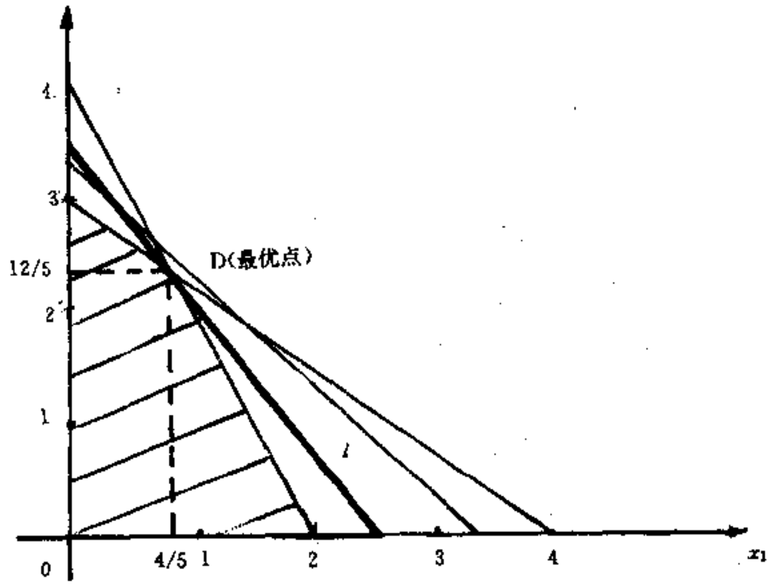


图 4-1

由于上图中 l 过可行域的顶点 D，且最优解为

$$x_1 = 4/5, x_2 = 12/5$$

目标函数的最大值为 $Z = 52/5$

这里我们也介绍一下集合，凸集的概念. 所谓集合是一些事物的整体. 如图 4-1 中阴影部分就是可行解的集合. 凸集是连结集合中任意两点间线段上的点仍属于集合. 在几何形体上看，凡是凸的形体，连结形体中任意两点的直线段，都落在形体之中. 图 4-1 中的阴影部分也组成一个凸集合.

例 1.1 所示的线性规划问题的最优解，取在其可行域的顶点. 可以证明线性规划的最优解一定在其凸可形域的顶点上取得.

五、解线性规划的单纯形法

线性规划问题的最优解产生在基本可行解之中，所以在求得线性规划的解时，不必研究所有的基本解. 因而在求解的过程中寻求最优解的范围就被大大的缩小了.

1. 单纯形法的求解过程

单纯形法寻找最优解不是一次完成的，实际上是个迭代的过程.当每次找到一个成功的基本可行解，都可以使目标函数比上次有所增加.而每次基本可行解的获得又是由一个非基变量代替一个基变量而完成的，新的基本变量的形成满足如下过程：首先，令一个不合适基本变量为零，再从非基本变量中划出一个合适的进入基本变量，计算是否满足(式4.4).基变量代换应满足下面的原则，一是新变量变成基本变量的组成部分之后，应使目标函数值有所增加.二是离开的一个基变量应保证其可行性不被破坏.这个迭代过程可以一直做下去，直到找到最优解为止.

下面用例1.1具体说明单纯形法的求解过程.

(1)将原始线性规划问题的约束上加上松驰变量 x_3、x_4、x_5，使问题化为标准型，即

$$\max \quad Z = 4x_1 + 3x_2$$

$$\text{约束} \quad \begin{aligned} 3x_1 + 4x_1 + 3x_3 &= 12 \\ 3x_1 + 3x_2 + x_4 &= 10 \\ 4x_1 + 2x_2 + x_5 &= 8 \end{aligned}$$

这里约束等式中右端的常数一定要大于或等于零.

(2)将上述问题改写为下面等价形式

$$Z - 4x_1 - 3x_2 = 0$$

$$\begin{aligned} 3x_1 + 4x_2 + x_3 &= 12 \\ 3x_1 + 3x_2 + x_4 &= 10 \\ 4x_1 + 2x_2 + x_5 &= 8 \end{aligned}$$

(3)选择初始基本可行解

把变量 x_1、x_2 置于零，即令 $x_1 = x_2 = 0$，则可得到另外三个变量的值，$x_3 = 12$，$x_4 = 10$，$x_5 = 8$ 目标函数值为 $Z = 0$.

在单纯形法的迭代过程中，要求每个约束方程都有这样一个变量：在该方程中这个变量的系数为1，在其它约束方程中(包括目标函数方程)其系数为零.由这些变量构成基变量，这时基的系数矩阵为单位阵.因此，基变量的数值容易求出.

(4)选择进基变量

选择进基变量的原则是：取目标函数中具有绝对值最大负系数的非基变量作为进基变量.

由目标函数 $2 - 4x_1 - 3x_2 = 0$ 可以看到 x_1 对目标函数有最大的增加量(4).因此，只有使具有负数系数的非基变量进基，目标函数才能得到改善，并且能减少迭代的次数.当非基变量的系数都大于或等于零，则说明已得到了最优解.

(5)选择离基变量

选择离基变量的原则是：用已选定的进基变量的一列系数，分别除各所在自行方程右端的常数，取最小非负比值所在行的基变量为离基变量.

在例1.1中

$$\frac{b_1}{a_{11}} = \frac{12}{3} = 4 \quad \frac{b_2}{a_{21}} = \frac{10}{3} = 3\frac{1}{3} \quad \frac{b_3}{a_{31}} = \frac{8}{4} = 2$$

因此，x_5 为基变量.

当 x_1 已经被选为进基变量时，x_5 被选为离基变量的依据是怎样体现的呢？

设 $x_2=0$，由
$$3x_1+4x_2+x_3=12$$
$$3x_1+3x_2+x_4=10$$
$$4x_1+2x_2+x_5=8$$
分别解出 x_3, x_4', x_5 与 x_1 的关系
$$x_3=12-3x_1 \quad x_4=10-3x_1 \quad x_5=8-4x_1$$

这里应保证：x_3、x_4、x_5 中有一个值为 0，另两个为正. 因此，当 x_3、x_4、x_5 分别为 0 时，有

$x_3=0$ 有 $x_1=4$

$x_4=0$ 有 $x_1=3\frac{1}{3}$

$x_5=0$ 有 $x_1=2$

我们也可推得：

$x_3=0$ 时，$x_4<0, x_5<0$

$x_4=0$ 时，$x_3<0, x_5<0$

$x_5=0$ 时，$x_3>0, x_4>0$

因此，取 x_3, x_4 为 0 时，得不到可行解，所以 x_5 具备了两个因素：

(a)进基 x_1 使目标函数的取值尽可能的大；

(b)新的基本解是可行的，即所有变量非负；

(6)换基，消元.

在选定的离基变量所在的方程中，将进基变量的系数变为 1，把其他方程中的进基变量系数变为零，这个过程就是换基，消元过程. 换基，消元的过程可用高斯消元法来实现，例 5-2 的换基、消元过程是

$$\begin{pmatrix} -4 & -3 & 0 & 0 & 0 & 0 \\ 3 & 4 & 1 & 0 & 0 & 12 \\ 3 & 3 & 0 & 1 & 0 & 10 \\ (4) & 2 & 0 & 0 & 1 & 8 \end{pmatrix} \sim \begin{pmatrix} -4 & -3 & 0 & 0 & 0 & 0 \\ 3 & 4 & 1 & 0 & 0 & 12 \\ 3 & 3 & 0 & 1 & 1 & 10 \\ 1 & \frac{1}{2} & 0 & 0 & \frac{1}{4} & 2 \end{pmatrix}$$

$$\sim \begin{pmatrix} 0 & -1 & 0 & 0 & 1 & 8 \\ 0 & (\frac{5}{2}) & 1 & 0 & -\frac{3}{4} & 6 \\ 0 & \frac{3}{2} & 0 & 1 & -\frac{3}{4} & 4 \\ 1 & \frac{1}{2} & 0 & 0 & \frac{1}{4} & 2 \end{pmatrix}$$

此时，$x_1=2, x_3=6, x_4=8, Z=8$，但是目标函数中仍有负系数.

(7)重复(4)～(6)步有

$$
\sim\begin{pmatrix} 0 & -1 & 0 & 0 & 1 & 8 \\ 0 & 1 & \frac{2}{5} & 0 & -\frac{3}{10} & \frac{12}{5} \\ 0 & \frac{3}{2} & 0 & 1 & -\frac{3}{4} & 4 \\ 1 & \frac{1}{2} & 0 & 0 & \frac{1}{4} & 2 \end{pmatrix}\sim\begin{pmatrix} 0 & 0 & \frac{2}{5} & 0 & \frac{7}{10} & \frac{52}{5} \\ 0 & 1 & \frac{2}{5} & 0 & -\frac{3}{10} & \frac{12}{5} \\ 0 & 0 & -\frac{3}{5} & 1 & -\frac{3}{10} & \frac{2}{5} \\ 1 & 0 & -\frac{1}{5} & 0 & \frac{2}{5} & \frac{4}{5} \end{pmatrix}
$$

此时，目标函数中没有负系数，且

$x_1=\frac{4}{5}\quad x_2=\frac{12}{5}\quad x_4=\frac{2}{5}\quad Z=\frac{52}{5}$

则等价问题的最优解为

$x_1=\frac{4}{5},x_2=\frac{12}{5},x_3=0,x_4=\frac{2}{5},x_5=0$

原始问题的解优解为

$x_1=\frac{4}{5},x_2=\frac{12}{5}$

原始目标函数的最大值为

$$
Z=4\times\frac{4}{5}+3\times\frac{12}{5}=\frac{52}{5}
$$

使用 Mathematica 软件可以很方便地求解各类线性规划问题．例如，我们可以用命令 Constrainedmax 来求若干个不等式约束下的极大值，用命令 Constrainedmin 来求若干个不等式约束下的极小值．

我们将例 1.1 运用 Mathematica 求解有：

```
In[1]=Constrainedmax[4x+3y,{3x+4y<=12,3x+3y<=10,4x+2y<=8},{x,y}]
```

Out[1]= $\{\frac{52}{5},\{x\to\frac{4}{5},y\to\frac{12}{5}\}\}$

2. 单纯形表

由上述的例子可以看出，线性规划问题的求解过程采用了高斯消去法的运算过程，其功能与解线性方程组的增广阵相类似．因此，我们可以设一张表格，称之为单纯形表，使单纯形法的迭达运算在这张表上进行．如例 5-2 的单纯形表表示如下．

初始单纯形表为：

其变量	x_1	x_2	x_3	x_4	x_5	解
Z	-4	-3	0	0	0	0
x_3	3	4	1	0	0	12
x_4	3	3	0	1	0	10
x_5	[4]	2	0	0	1	8

迭代一次后，有单纯形表：

基变量	x_1	x_2	x_3	x_4	x_5	解
Z	0	-1	0	0	1	8
x_3	0	$\left[\frac{5}{2}\right]$	1	0	$-\frac{3}{4}$	6
x_4	0	$\frac{3}{2}$	0	1	$-\frac{3}{4}$	4
x_1	1	$\frac{1}{2}$	0	0	$\frac{1}{4}$	2

经过第二交迭代后，有单纯形表：

基向量	x_1	x_2	x_3	x_4	x_5	解
Z	0	0	$\frac{2}{5}$	0	$\frac{7}{10}$	$\frac{52}{5}$
x_2	0	1	$\frac{2}{5}$	0	$-\frac{3}{10}$	$\frac{12}{5}$
x_4	0	0	$-\frac{3}{5}$	1	$-\frac{3}{10}$	$\frac{2}{5}$
x_1	1	0	$-\frac{1}{5}$	0	$\frac{2}{5}$	$\frac{4}{5}$

3. 单纯形法的扩展

(1). 人工变量

线性规划一般问题中，可以有“＝”或“≥”的约束. 当这种情况发生时，松弛变量的系数不能构成单位阵. 因此，找不到初始基本解. 为解决这一问题，我们可以采用加人工变量的方法，使问题可以用单纯形法来解决.

例 1.3 解下述线性规划问题

$$\max \quad Z = 3x_1 - x_2 - x_3$$

$$\text{约束} \quad x_1 - 2x_2 + 3x_3 \leqslant 11$$

$$-4x_1 + x_2 + 2x_3 \geqslant 3$$

$$-2x_1 + x_3 = 1$$

在约束条件中加入松弛变量，剩余变量和工人变量，可得下面的标准型

$$\max \quad Z = 3x_1 - x_2 - x_3$$

$$\text{约束} \quad x_1 - 2x_2 + x_3 + x_4 = 11$$

$$-4x_1 + x_2 + 2x_3 - x_5 + x_6 = 3$$

$$-2x_1 + x_3 + x_7 = 1$$

$$x_i \geqslant 0 \quad (i = 1,2,\cdots,7)$$

这里的 x_6, x_7 为人工变量.

由于 x_6, x_7 是加入原约束方程组中的虚拟变量，因此，在求解带有人工变量的问题中，应先把人工变量从基变量中替换出去. 如果经过基变换，基变量中不再含有人工变量，则原问题的解存在；反之，原问题无解.

(2). 两阶段法

两阶段法是将增加了人工变量的线性规划问题分成两个阶段来求解.

第一阶段就是判断原线性规划问题是否存在基本可行解. 也就是设法把人工变量替换出基变量，使寻找到的基本可行解中不含人工变量. 其具体方法是构造所有人工变量的和做为辅助函数，设为

$$\min \quad \bar{Z} = x_{n+1} + x_{n+2} + \cdots + x_{n+l}$$

求此目标函数在原问题约束条件下的最优解，并使 $\bar{Z}=0$，则人工变量就都是非基本变量了，这时得到的解是不含人工变量的基本可行解，从而完成了第一阶段的任务.

第二阶段的求解是在第一阶段的基础上进行的. 把第一阶段得到的不含人工变量的基本可行解做为始基本可行解，运用单纯形法继续迭代，求得最优解.

对例 1.3 我们可以采用两阶段法求其最优解.

设辅助目标函数为

$$\min \quad \bar{Z} = x_6 + x_9$$

将目标函数化为标准型，有

$$\max \quad Z' = -\bar{Z} = -x_6 - x_7$$

因此，第一阶段的初始单纯形表为

基向量	x_1	x_2	x_3	x_4	x_5	x_6	x_7	解
Z'	0	0	0	0	0	1	1	0
Z	-3	1	1	0	0	0	0	0
x_4	1	-2	1	1	0	0	0	11
x_6	-4	1	2	0	-1	1	0	3
x_7	-2	0	1	0	0	0	1	1

首先消去辅助函数方程中，x_6、x_7 的系数，并选定 x_3 为进基变量，x_7 为离基变量. 其单纯形表为：

解向量	x_1	x_2	x_3	x_4	x_5	x_6	x_7	解
Z'	6	-1	-3	0	1	0	0	-4
Z	-3	1	1	0	0	0	0	0
x_4	1	-2	1	1	0	0	0	11
x_6	-4	1	2	0	-1	1	0	3
x_7	-2	0	(1)	0	0	0	1	1

利用单纯形法进行迭代，直到使目标函数达到最优，第一次迭代后，有单纯形表：

基向量	x_1	x_2	x_3	x_4	x_5	x_6	x_7	解
Z'	0	−1	0	0	1	0	3	−1
Z	−1	1	0	0	0	0	−1	−1
x_4	3	−2	0	1	0	0	−1	10
x_5	0	(1)	0	0	−1	1	−2	1
x_3	−2	0	1	0	0	0	1	1

进行第二交迭代后，有单纯形表：

解向量	x_1	x_2	x_3	x_4	x_5	x_6	x_7	解
Z'	0	0	0	0	0	1	1	0
Z	−1	0	0	0	1	−1	1	−2
x_4	(3)	0	0	1	−2	2	−5	12
x_2	0	1	0	0	−1	1	−2	1
x_3	−2	0	1	0	0	0	1	1

这时，由单纯形表中可以看到，辅助目标函数已经为 0，则第一阶段的运算结束. 由于不含有人工变量的基本解已经找到，所以在第二阶段运算中，辅助目标函数和人工变量已无用处，在前面第二次迭代后得到的单纯形表中，可以将它们去掉，并在其基础上再次迭代，有最后的单纯形表：

基向量	x_1	x_2	x_3	x_4	x_5	解
Z	0	0	0	$\frac{1}{3}$	$\frac{1}{3}$	2
x_1	1	0	0	$\frac{1}{3}$	$-\frac{2}{3}$	4
x_2	0	1	0	0	−1	1
x_3	0	0	1	$\frac{2}{3}$	$-\frac{4}{3}$	9

因此，原线性规划问题的最优解为：

$$x_1 = 4 \quad x_2 = 1 \quad x_3 = 9$$

松弛变量 $x_4=0$

剩余变量 $x_5=0$

最优目标函数值 $Z=2$.

4. 单纯形法小结

①准备工作

将一般线性规划问题化为标准形.

②第一阶段的运算

(a) 构造辅助目标函数，即所有人工变量的和，并求其最优解.

(b) 当辅助目标函数值为 0,则基变量中不包含人工变量,可转入第二阶段的运算.如果此时辅助目标函数值不为 0,则基变量中仍含有人工变量,此时原始问题无可行解,运算到此停止.

③第二阶段的运算

在第一阶段运算的基础上,去掉辅助目标函数所在行与人工变量所在列,对原目标函数进行单纯形法迭代,直到求得线性规划的最优解.

六、线性规划的对偶问题及求解原理

1. 对偶定义:若有线性规划问题

$$\max \quad Z=CX$$
$$AX\leqslant b$$
$$X\geqslant 0$$

则有线性规划问题

$$\min \quad Z=\lambda b$$
$$\lambda A=C$$
$$\lambda\geqslant 0$$

与之相对应,后一个线性规划叫做前一个问题的对偶.

如例 1.1 的对偶问题为:

$$\min \quad Z'=12\lambda_1+10\lambda_2+8\lambda_3$$

约束 $3\lambda_1+3\lambda 2+4\lambda_3\geqslant 4$

$$4\lambda_1+3\lambda_2+2\lambda_3\geqslant 3$$
$$\lambda_1,\lambda_2,\lambda_3\geqslant 0$$

由此我们可以看到,线性规划的对偶问题有以下的特点:

(1)线性规划问题都对对偶问题,并且形式上对称.若原问题求最大值,则对偶问题是求最小值;两个问题约束的系数阵互为转置;目标函数的系数与约束右端常数互换位置;最优函数值相等.

(2)对偶问题的对偶是原始问题.

2. 对偶问题的单纯形算法步骤

(1)化对偶问题为标准型.

(2)如对偶约束右端常数为正时,可用前面的解法计算.如对偶问题中约束右端常数为负时.首先,选择约束右端常数中具有最小负数值的约束方程中,原来的基变量为离基变量.取目标函数中非基变量的系数与离基变量所在的约束条件中相应系数的最大负比值对应的非变量为进基变量.其次,采用消元、换基完成迭代,一直到约束右端的常数全部为正时,则停止运算,此问题取得最优解.

第二节　非线性规划模型

一、资源分配

例 2.1　在两个过程中,在总资源一定的条件下,如何使输出总量最大.

分析：假设

$f(x_1)$——过程Ⅰ的输出量

$f(x_2)$——过程Ⅱ的输出量

A——资源总量

x_1——过程消耗的资源

x_2——过程消耗的资源

由此，可以建立这一问题的数学模型

$$\max \quad Z=f(x_1)+f(x_2)$$

$$约束 \quad x_1+x_2=A$$

二、最佳定货量

例 2.2 当某仓库每次订货的费用和每天的存费为常数，在库存用尽时立即进货，且这一消耗、补充的过程可周期地进行．怎样确定最佳货量，而使费用最少．

分析：假设

a、b、c——每次定货费和每天的储存费（存货价值百分比），每件货物的价值．

x——一次进货的数量．

Q——设定时间内的定货总量．

Z——设定时间内所需总费用．

由于问题中，仓库中消耗零时，可以立即得到补充，则库存为不允许缺货．因此可知如下的概念：

$x/2$——平均库存量．

Q/x——定货次数

有 $Z=x/2\cdot bc+Q/x\cdot a$

求最佳定货量，而使总费用最少，即是求

$$\min \quad Z=\frac{bc}{2}\cdot x+aQ\cdot\frac{1}{x}$$

解：由于目标函数为一元函数，当

$$\frac{dz}{dx}=\frac{bc}{2}+aQ\cdot\frac{-1}{x^2}=0 \text{ 时，}$$

可解得 $x_0=\sqrt{2aQ/bc}$ 且

$$\left.\frac{d^2T}{dx^2}\right|_{x_0}=\left.\frac{2aQ}{x^3}\right|_{x_0}>0$$

因此，总费用在 $x_0=\sqrt{2aQ/bc}$ 时最低，并且为 $Z=\sqrt{2abcQ}$

当 $Q=8\,000$ 件，$a=12.50$ 元，$b=20\%$，$c=1$ 元，时

$$2=0.1x+\frac{100\,000}{x}$$

三、非线性规划的概念

对于例 2.1、例 2.2 中的目标函数关于 x_1、x_2 不是线性函数，例 2.2 中的目标函数亦关于 x 不是线性函数．因此，我们称目标函数是非线性函数或条件是非线性函数的数学

规划问题称为非线性规划问题.

1. 非线性规划的数学模型

形式为：

$$\min \quad Z = f(X) \tag{4.6}$$

$$约束 \quad h_i(X) = 0 \qquad i = 1,2,\cdots,m \tag{4.7}$$

$$g_j(X) \geqslant 0 \qquad j = 1,2,\cdots,l \tag{4.8}$$

其中 $X=(x_1,x_2,\cdots,x_n)^T$ 是 n 维间的点.

我们可以采用在目标指数两端以 -1，使 $\max \cdot f(X)$ 化为 $\min -f(X)$，以 -1 乘不等式约束，使"$\leqslant$"化为"$\geqslant$".

2. 非线性规划问题的类型

(1)无约束问题

在前面已给的非线性规划的数学模型中，如果没有约束式(式 4.7)和(式 4.8)时，即为无约束问题.

无约束问题在求最小值时，可直接利用高等数学课程中求极值的方法，如 2.2，也可以利用众多的搜索法(或称为迭代法)，在计算机上实现运算.

(2)有约束问题

在一般的非线性规划问题中，约束条件极其重要，它决定着最优解是否可行的问题.

通常采用高等数学课程中的消元法或拉格朗日乘数法，将其化为无约束问题来求解.也可以采用将非线性问题线性化的方法来求解.

总之非线性规划的求解方法比较复杂，我们在这本书中不做详细地介绍，对多元无约束问题给出一个用 BASIC 语言编写的通用梯度法程序，把复杂的运算交给计算机来完成(具体的运算理论，可参考清华大学出版社出版的《运筹学》).

四、解无约束问题的最速下降法及实验

1. 最速下降法的基本思想

在无约束问题 $\min \quad Z=f(X)$ 中，设是 X 是 n 维空间 E^n 的点.

如果 $X^k \in E^n$ 是给定的初始点，沿使 $f(X)$ 下降最快的负梯度方向.

$$P^k = -\nabla f(X^k) = -\left(\frac{\partial f}{\partial x_1},\frac{\partial f}{\partial x_2},\cdots,\frac{\partial f}{\partial x_n}\right)\Big|_{X^k}$$

并取步长 λ_k 进行迭代，使

$$X_{k+1} = X_k + \lambda_k P_k$$

满足 $f(X_{k+1})<f(X_k)$，λ_k 就成为最优步长.

2. 最速下降法的迭代步骤

(1). 给定初始点 $x_0 \in E^n$. 精度 $\varepsilon>0$，并且令 $k=0$；

(2). 计算 $\nabla f(X^k)$；

(3). 若 $\|\nabla f(X^k)\| \leqslant \varepsilon$，则停止，得 $X^*=X^k$(X^* 为最优解)，否则转到(4)；

(4). 令 $P^k=-\nabla f(X^k)$ 求

$$\min f(X^k+\lambda_k P^k)=f(X^k+\lambda_k P^k)$$

(5). 令 $X^{k+1}=X^k+\lambda_k P^k$，返回步骤(2).

3. 最速下降法的 BASIC 程序

在程序中可任意调换目标函数式，对例 2.2 当 $x_0=10$，计算的精度为 0.001 时的计算结果 $x=1\ 000.143$ 而最省费用是 200 元.

```
100  REM  Method of steeepest descent
110  CLS
120  PINT"已知":
130  INPUT"        1. 自变量个数=";N
140  PRINT"        2. 自变量初值="
150  DIM X(N),F9(N),Z(N),A(N)
160  FOR I=1 TO N:LOCATE 3,8+I*14:PRINT "X";I;:INPUT X(I):NEXT
     I
170  INPUT"        3. 计算精度=";E:PRINT
180  S=0:H1=.001
190  X(I)=X(I)+H1:GOSUB 410:P1=F
200  X(I)=X(I)-2*H1:GOSUB 410:P2=F
210  X(I)=X(I)-2*H1:GOSUB 410:P2=F
220  F9(I)=(P1-P2)/(2*H1):S=S+F9(I)*F9(I)
230  X(I)=X(I)+H1
240  NEXT I
250  S=SQR(S)
260  IF S<E THEN 370
270  L=0:H2=.1
280  FOR I=1 TO N:Z(I)=X(T):NEXT I:GOSUB 410:FO=F
290  FOR I=1 TO N:X(I)=Z(I)-(L+H2)*F9(I):NEXT I:GOSUB 410:F1=
300  IF F1<F0 THEN 360
310  IF ABS(H2)<E THEN 360
320  IF L=0 THEN 340
330  H2=-H2/4:GOTO 290
340  H2=H2/4:GOTO 290
350  L=L+H2:H2=2*H2:F0=F1:FOR I=1 TO N:A(I)=X(I):NEXT I:GO-
     TO 290
360  FOR I=1 TO N:X(I)=A(I):NEXT I:GOTO 180
370  PRINT"优化结果:"
380  GOSUB 400:PRINT TAB(10);"F(x)min=";F
390  FOR I=1 TO N:PRINT TAB(13);"X";I;"=";X(I):NEXT I
```

```
400  END
410  F=.2/2*X(1)+12.5*80 000/X(1)
420  RETURN
RUN
已知：
1. 自变量个数=? 1
2. 自变量初值=X1? 10
3. 计算精度=? 0.001
优化结果：
F(x)min=200
X1=1000.143
Ok
```

五、解有约束问题的外罚函数法及实验

1. 罚函数法的基本思想

罚函数法是化有约束为无约最优化的方法之一，它利用目标函数和约束条件构造一个新的函数，而得到的最优解是原问题的最优解.

外罚函数的迭代点一般在其可行域外部移动，当惩罚逐渐加大时，可以使迭代点靠近可行域.

对 min $Z=f(X)$，在不等式约束 $g_j(X)\geqslant 0(j=1,2,\cdots,l)$ 条件下的非线性规划，取罚函数为：

$$T(X,M_k)=f(X)+M_k\sum_{j=1}^{l}\{\min[0,g_j(X)]\}^2$$

其中 $0<M_1<M_2<M_3<\cdots<M_k<M_{k+1}<\cdots$

$$\lim_{k\to\infty}M_k=+\infty$$

并且 $\min[0,g_j(X)]=\begin{cases}g_j(X) & 若\ g_j(X)<0\\ 0 & 若\ g_j(X)\geqslant 0\end{cases}$

罚互数 $T(X,M_k)$ 无约束极值问题的最优解 $X^k=X^k(M_k)$ 看做是以 M_k 为参数的一条轨迹，并且当取 $0<M_1<M_2<\cdots<M_k<M_{k+1}<\cdots,M_k\to\infty$ 时，点到 $\{X(M_k)\}$ 沿着这条轨迹趋于条件极值的最优解，即原问题的最优解 X^*.

2. 迭代步骤

(1)取 $M_1>0$(例如取 $M_1=0.5$ 或 1)，允许误差 $e>0$，令计算次数 $k=1$.

(2)求无约束条件条件极值问题的最优解 $X^k=X(M_k)$：

$$T(X^k,M_k)=\min T(X,M_k)=f(X)+M_k\sum_{j=1}^{l}\{\min[0,g_{j1}(X)]\}^2$$

(3)检验 $-g_j(X^k)\leqslant e(j=1,2,\cdots,l)$. 若条件满足则 $X^*=X^k$ 为最优解；反之，取 $M_{k+1}>M_k$，令 $k=k+1$，转到步骤(2).

3. 外罚函数法的 BASIC 程序

本程序可以任意调换目标函数和约束条件，通用性较强.
在程序请单后面附有下面二次规划的解.

$$\min Z = x_1^2 + x_2^2 - 8x_1 - 10x_2$$

$$约束 \quad -3x_1 - 2x_2 + 6 \geqslant 0$$

$$x_1 \geqslant 0 \quad x_2 \geqslant 0$$

```
100  REM Sequential unconstrained minimization technique
110  CLS
120  PRINT"已知:"
130  INPUT "    1. 自变量个数      =";N
140  INPUT "    2. 等式约束个数    =";L
150  INPUT "    3. 不等式约束个数=";M
160  DIM X(N),Y(N),A(N),B(N),G(L+M)
170  PRINT "    4. 自变量初值      ="
180  FOR I=1 TO N:LOCATE 5,8+I*14:PRINT "X";I;:INPUT X(I):NEXT
     I
190  INPUT "    5. 计算精度        =";ESP:PRINT
200  CM=1:CO=4
210  GOSUB 300
220  REM PRINT "EPS=";EPS;">=";ABS(RM);"?"
230  IF ABS(RM)<=EPS THEN 250
240  CM=CO*CM:GOTO 210
250  PRINT "优化结果:"
260  PRINT "    F(x)min=";F
270  FOR I=1 TO N:PRINT"X";I;"=";X(I):NEXT I
280  FOR I=1 TO L+M:PRINT "G";I;"=";G(I):NEXT I
290  END
300  D=1:Z=1:E=.001:Q=1:REM Hook-Jeeve
310  J=1
320  FOR I=1 TO N:X(I)=Y(I):NEXT I
330  GOSUB 600:GOSUB 560:F1=PX
340  X(J)=Y(J)+D*Z:GOSUB 600:GOSU 560:F2=PX
350  IF F2<F1 THEN 380
360  X(J)=Y(J)-D*Z:GOSUB 600:GOSU 560:F2=PX
370  IF F2<F1 THEN 380 ELSE 390
380  FOR I-1 TO N:Y(I)=X(I):NEXT I
390  IF J<N THEN 520
400  FOR I=1 TO N:X(I)=A(I):NEXT I
```

```
410  GOSUB 600:GOSUB 560:F2=PX
420  FOR I=1 TO N:X(I)=Y(I):NEXT I
430  GOSUB 600:GOSUB 560:F2=PX
440  IF F2<F1 THEN 480
450  IF D<=E THEN 530
460  FOR I=1 TO N:Y(I)=A(I):ENXT I
470  D=D/2:GOTO 310
480  FOR I=1 TO N
490  B(I)=Y(I):Y(I)=B(I)+Q*(B(I)-A(I)):A(I)=B(I)
500  NEXT I
510  GOTO 310
520  J=J+1:GOTO 320
530  FOR I=1 TO N:A(I)=X(I):NEXT I
540  PX=F1
550  RETURN
560  PH=0:FOR I=1 TO L:PH=PH+G(I)*G(I):NEXT I
570  PQ=0:FOR I=L+1 TO L+M:PQ=PQ+((ABS(G(I))-G(I))/2)^2:
     NEXT I
580  RM=CM*(PH+PQ):PX=F+RM
590  RETURN
600  F=X(1)^2+X(2)^2+8*X(1)-10*X(2)
610  G(1)=-3X*X(1)-2*X(2)+6
620  G(2)=X(1)
630  G(3)=X(2)
640  RETURN
RUN
已知:
1.自变量个数      =? 2
2.等式约束个数    =? 0
3.不等式约束个数=? 3
4.自变量初值      =X1? 0
                   X2? 0
5.计算精度        =? 0 001
优化结果:
F(X)min=-2 130 634
X1=.328 125
X2=250 713
G1=0
G2=.328 125
```

$G3=2.507\ 813$

Ok

第三节 动态规划模型

动态规划是解决多阶段最优决策过程的一种数学方法，其基本思想是将问题按一定的原则，分成若干个可以控制的阶段，并对各个阶段依次变成最优，从而达到整体最优.

应当指出的是，动态规划没有标准的数学算法（如前面线性规划的单纯形和非线性规划的罚函数法等.）它只是解决一类问题的方法或途径，在运用动态规划是建立模型时，应在正确理解和运用其基本概念、基本原理的基础上，发挥想象力、创造力建立和求解模型.多阶段的决策过程中，依次采取的一系列最优的决策，使其控制过程的发展，这个决策可以用数量来表示，当这一数量指标达到最大或最小时，过程在此阶段也达到了优，如果一个过程的阶段可用时段来划分，相互联系的若干时段组成了整个多阶段决策过程的数学规划问题，并具有其中一个阶段的决策影响到下一个阶段的决策的特点.

当在静态最优化问题中引入时段的因素后，也可以用动态规划的方法来求得最优解.

一、管道铺设线路的最优化问题

1.实例

例 3.1　从地点 A 铺设一条天然气管道到地点 E，如图 4-2，中间必须经进三个中间站，第一站可以在 $\{B_1、B_2、B_3\}$ 中任选一地点.第二、三站可分别在地点 $\{C_1、C_2、C_3\}$ 和 $\{D_1、D_2\}$ 中选择.两点间的连线表示有管道连接，连线上的数字表示铺本段管道所需费用.而当两点间没有连线，则表示该两点根本不能铺设管道，怎样选择从 A 到 E 的铺设路线，使总费用最省.

2.线路铺设的最优原则和基本方法

(1)最优原则

在线路铺设的过程中，如果一条线路的铺设费用最省，则对此线路上的任何中间站来说，以此中间站为起点直到终点的线路铺设费用最省.

(2)基本方法和步骤

根据上述的最优原则，可以先把定过程分为四个阶段，最后一个阶段 $D\to E$；倒数第二个阶段 $C\to D$；第二个阶段 $B\to C$；第一个阶段 $A\to B$.然后从（图 4-2）中沿 E 向 A 的方向分阶段决策，寻求最佳的策略，最终形成一个最佳的铺设路线.

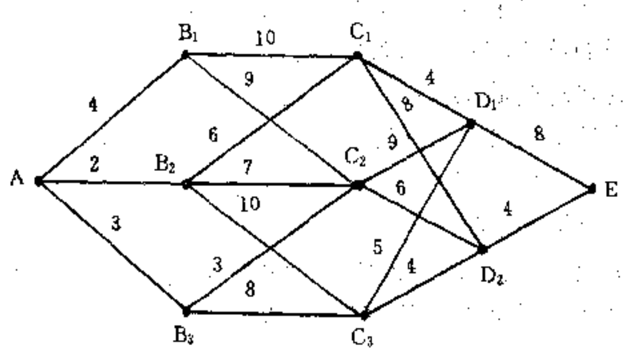


图 4-2

第一步：从最后一个阶段开始，寻求 $D\to E$ 的最一佳路线，用 $f_i(i=1,2,3,4)$，表示第 i 阶段的最省费用.

$$f_4(D_2)=\min\begin{Bmatrix}(D,E)=8\\(D_2,E)=4\end{Bmatrix}=4$$

第二步：寻求 $C\to E$ 的最佳路线，此时

$$f_3(C_3)=\min\begin{Bmatrix}(C_1,D_2)+f_4(D_2)=8+4\\(C_2,D_2)+f_4(D_2)=6+4\\(C_3,D_2)+f_4(D_2)=4+4\end{Bmatrix}=8$$

第三步：寻求 $B\to E$ 的最佳路线，此时

$$f_2(B_3)=\min\begin{Bmatrix}(B_2,C_3)+f_3(C_3)=10+8\\(B_3,C_3)+f_3(C_3)=8+8\end{Bmatrix}=16$$

第四步：同上面的分析过程一样，寻求 $A\to E$ 的最佳路线：

$$f_1(A)=\min\{(A,B_3)+f_2(B_3)=3+16\}=19$$

由此可以得到最佳中路线 $A\to B_3\to C_3\to D_2\to E$，总费用为 19 个单位.

根据第一步至第四步的决策可以发现，我们寻求最佳路线的步骤方向与 AE 的方向相反. 在每一阶段的决策中都能选阶段的中间站直到终点的 E 的最佳路线，每一步骤的决策中仅用到上一步骤的结果. 因此，可以把上述的求解过程中所用到的方法称为迭代法，并可在计算机上实现运算，找到最佳路线.

二、动态规划问题的基本概念、原理和逆推方程

1. 基本概念

(1)阶段

将所给的问题恰当地分成若干相互联系、互不交叉的阶段，每一阶段的求解相对方便. 一般用 k 描述阶段变量.

(2)状态及状态变量

状态是某阶段的初始位置，同时也是前一阶段的终点，各个阶段的状态演变构成了多阶段的终点. 各个阶段的状态演变成了多阶段决策过程. 若用 x_k 表示阶段的状态，则 x_k 称为状态变量. 状态变量可以是一个变量，也可以是一组变量的集合，因此，状态变量的个数也反映出状态的维数情况.

(3)决策及决策变量

决策是某阶段状态给定后，从该状态演变到下一个阶段某一状态的选择，如例 3.1 中的 D_2 已经被最后阶段选择，如何寻找 C 使 $C\to E$ 成为最佳线路的选择，我们选择了 C_3，也就做出了决策，可用 $U_k(x_k)$表示第 k 阶段，在 x_k 状态下所做的决策，并称之为决策变量.

(4)策略与子策略

如果在多阶段决策过程中，决策变量 $U_k(x_k)(k=1,2,\cdots,n)$都已确定，则函数序列

$$\{U_1(x_1),U_2(x_2),\cdots,U_n(x_n)\}$$

称为过程的一个策略. 例 3.1 中 $A\to B_3\to C_3\to D_2\to E$ 就问题的一个策略，在实际应用中，策略加限制构成允许策略，而最优策略一定是允许策略.

从策略中某一阶段开始到终点的决策集合序列称为原策略的子策略，如例 3.1 中 $C_3\to D_2\to E$. 一般来说决策函数序列$\{U_n(x_n)|n=1,2,\cdots,n\}$的任一子序列都可称之为多阶

段决策过程中某段过程所对应的子策略.

(5)状态转移方程

在多阶段决策过程中,在第 k 阶段的决策变量 $U_k(x_k)$ 已经给定,则第 $k+1$ 阶段的状态变量就完全给定,即 x_{k+1} 为已知确定的量. 因此,x_{k+1} 取决于 x_k 和 U_k,则可记为 $x_{k+1}=T_k(x_k,U_k(x_k))$,表示了第 k 阶段到第 $k+1$ 阶段的状态转移规律,称之为转移方程.

(6)目标函数(指标函数)

在多阶段决策问题中,目标函数是用来衡量实现过程优劣的一个数量指标,是在给定阶段的前提条件下,该阶段的状态变量及以后各阶段的决策变量的函数,一般记目标函数为

$$V_{kl}(x_k,U_k,x_{k+1},U_{k+1},\cdots,x_{k+l-1},U_{k+l-1})$$

此式表示了由阶段 k 到阶段 l 各个状态和所包含的一系列决策.

目标函数的形式一般取"和"式或"积"式.

"和"式表达式:

$$V_{kl}=\sum_{j=k}^{l}V_j(x_j,u_j(x_j))$$

"积"式表达式:

$$V_{kl}=\prod_{j=k}^{l}V(x_j,U_j(x_j))$$

当 k 阶段的初始状态 x_k 给定时,指标函数是策略的确定函数,记

$$f_k(x_k)=\mathrm{opt}V_{kl}(x_k,U_k,x_{k+1},U_{k+1},\cdots,U_{k+l-1})$$

是指标函数的最优值. 在实际问题中 opt 分别表示 max 和 min.

2. 原理和递推方程

(1)最优化原理

一个过程的最优策略具有以下性质:即无论初始状态和初始决策怎样,其后各决策对以前的一个决策所形成的状态作为初始状态的过程而言,是最优策略,也就是说:最优策略的任一子策略必为最优策略.

(2)递推方程

由最优化原理,我们可得知,最优化策略可通过逐个阶段优化来实现. 因此,在实际问题的求解过程中可以推出解动态规划的基本方程:

$$f_k(x_k)=\underset{U_k\in D_k(x_k)}{\mathrm{opt}}\varphi\{V_k(x_k,U_k),f_{k+1}[T_k(x_k,U_k)]\}$$

其中 $x_k\in X_k$,X_k 为第 k 阶段可达状态集合,$D_k(x_k)$ 为第 k 阶段的允许策略集合.

推理方程显示了求最优函数的顺序与多阶段决策分段顺序是相反的,当初始条件是 $f_n(x_n)=0$ 时,可由 f_n 递推 f_{n-1},代入递推方程,可递推 f_{n-2},…以此类推,可求至 f_0. 即是最优函数. 当 $k\to\infty$ 时,最优函数以泛函方程的形式给出:

$$f(x)=\underset{u\in D}{\mathrm{opt}}\varphi\{V(x,u)f[T(x,u)]\}$$

3. 动态模型建立的步骤

运用动态规划的方法解决实际问题时,可按下面的步骤进行:

(1)将过程进行恰当的分段,并在分段时作出初始的决策.

(2)选择状态变量 x,使 x 既能反映过程演变的特征,又能满足无后效性,还能通过直

接或间接的方法得知它在各阶段的变量值，如例 3.1 中各阶段中间站间的铺设费用单位.

无后效性是指当某阶段状态给定后，以后的过程不受以前各阶段状态的影响.

(3)确定决策变量及每阶段的允许决策集合 $D_k(x_k)$.

(4)根据实际问题的变化过程机理写出状态转移方程：$x_{k+1}=T(x_k,u(x_k))$.

(5)写出目标函数(指标函数).

(6)写出最优函数的递推方程及初始和边界条件.过程阶段无限时，最优函数可写出为泛函方程的形式.

采用上面的方法和步骤，可对例 3.1 建立管道铺设的最省费用模型.

解：将由 $A\to E$ 的管道铺设分成四个阶段：

第一阶段 $A\to B$，第二阶段 $B\to C$，第三阶段 $C\to D$，第四阶段 $D\to E$.

令 x_k 表示第 k 阶段的状态变量；

u_k 表示第 k 阶段的决策变量；

f_k 表示第 k 阶段状态为 x_k 的最优值(最优费用).

$d_{k,k+1}$ 表示第 k 阶段到第 $k+1$ 阶段的费用可能值；

因此在 k 阶段内．$f_k(x_k)=\min\limits_{x_k}[U_k(x_k)]$

且状态转移方程为：

$$x_{k+1}=d_{k,k+1}+f_k(x_k)$$

令 f_k 表示第 k 阶段状态为 x_k 的最优值，则有动态模型的递推方程为：

$$f_k(x_k)=\min_{U_k}\{d_{kk+1}+f_{k+1}(x_{k+1}\}\quad k=1,2,3,4$$

其中所有的决策变量取自允许策略的集合.假设有初始条件 $f_5(x_5)=0$，至此，此模型建立完毕.

三、资源分配模型

设某种资源的总量为 a，分配给几个使用者，如何分配才能使总的收益最大？

解：将资源分配给一个或几个使用者的过程作为一个阶段，且

x_k——表示分配给第 k 个使用者的资源.

U_k——表示分配给第 k 个至第 n 个使用者的资源.

f_k——表示 U_k 条件下获得的最佳效益.

因此，状态转移方程为

$$U_{k+1}=U_k-x_k\quad k=1,2,\cdots,n$$

设 $h_k(x_k)$表示第 k 个使用者用 x_k 获得的效益.

x_k^* 表示 U_k 条件下的最佳决策

因此有

$$f_k(U_k,x_k^*)=\max\{h_k(x_k)+f_{k+1}(u_k-x_k,x_k^*\}$$

其中 u_k 取自允许策略集合，且设初始条件 $f_{k+1}(u_{k+1},x_{k+1}^*)=0$.

例 3.2　设有 8 个单位的总资源，欲分配三个使用者，每个使用者所获得的投资为 0～8 中任一整数，若每个使用者所获效益是同一函数，如何分配总资源，而使所获的效益最大？

表 4-2 给出了资源的不同分配情况下，所获效益的情况.

表 4-2

x	0	1	2	3	4	5	6	7	8
$g_1(x)$	0	5	15	40	80	90	95	98	100
$g_2(x)$	0	5	15	40	60	70	73	74	75
$g_3(x)$	0	4	26	40	45	50	51	52	53

上述的资源分配问题应如下假设：

(1). 每个分配方案获益与另一个方案的分配无关.

(2). 假设资源单位与获益单位不同，但所有方案的获益用同一单位度量.

(3). 总的获益额是每个资源使用者获益额的和.

解：设

x——表示使用者所用的资源数量.

$g_i(x)$——表示第 i 个使用者用 x 的获益额.

$d_i(x)$——表示第 i 个使用者为取得最大收益应得到的资源量.

$f_i(x)$——表示用 x 全部投入到第 i 个使用者的最大获益额.

我们采用动态规划的方法. 把总问题分成三个，并逐步将子问题解出，从而求得一定资源投入的最大收益，其步骤如下：

第一步：设只有第三个资源使用者可以获益，这时第三个使用者获益最大时，得到的资源量为：

$$d_3(x) = x \qquad (x = 0,1,2,\cdots 8)$$

最大获益为：

$$f_3(x) = g_3(x) \qquad (x = 0,1,2,\cdots 8)$$

我们可查表 4-2 得 x 不同时 $f_3(x)$ 及 $d_3(x)$ 值.

表 4-3

x	0	1	2	3	4	5	6	7	8
$f_3(x)$	0	4	26	40	45	50	51	52	53
$d_3(x)$	0	1	2	3	4	5	6	7	8

第二步：若有两个资源使用者，则用

$f_2(x)$——表示 x 个单位的资源分配给使用者三、二时所得的最大获益.

y——表示分配到第二个使用者的资源数量.

则 $x-y$——表示分配到第三个使用者的资源数量，因此有

$$f_2(x) = \max\{g_2(y) + f_3(x-y)\}$$

其中 $y=1,2,\cdots,x$.

$d_2(x)=y \quad (x=0,1,2,\cdots,8)$，且满足 $f_2(x)$.

因此，我们有：

$$x=0\text{ 时}\quad f_2(0)=\max_{y=0}\{g_2(y)+f_3(x-y)\}$$
$$=g_2(0)+f_3(0)=0$$

$$d_2(0)=0$$

$$x=1\text{ 时}\quad f_2(1)=\max_{y=0,1}\{g_2(y)+f_3(x-y)\}$$
$$=\max\begin{Bmatrix}g_2(0)+f_3(1)\\g_2(1)+f_3(0)\end{Bmatrix}$$
$$=\max\begin{Bmatrix}0+4\\5+0\end{Bmatrix}=5$$
$$d_2(1)=1$$

$$x=2\text{ 时}\quad f_2(2)=\max_{y=0,1,2}\{g_2(y)+f_3(x-y)\}$$
$$=\max\begin{Bmatrix}g_2(0)+f_3(2)\\g_2(1)+f_3(0)\\g_2(2)+f_3(0)\end{Bmatrix}$$
$$=\max\begin{Bmatrix}0+6\\5+4\\15+0\end{Bmatrix}=26$$
$$d_2(2)=0$$

由上式可知，当 $d_2(2)=0$ 时，表示第三使用资源者把资源 x 全部占有，第二个使用者不占有资源.

$$x=3\text{ 时}\quad f_2(3)=\max_{y=0,1,2,3}\{g_2(y)+f_3(x-y)\}$$
$$=\max\begin{Bmatrix}g_2(0)+f_3(3)\\g_2(1)+f_3(2)\\g_2(2)+f_3(1)\\g_2(3)+f_3(0)\end{Bmatrix}$$
$$=\max\begin{Bmatrix}0+40\\5+26\\15+4\\40+0\end{Bmatrix}=40$$
$$d_2(3)=0\text{ 或 }d_2(3)=3$$

依此类推，可得第一步和第二步的总结果，列表 4-4

表 4-4

x	0	1	2	3	4	5	6	7	8
$f_3(x)$	0	4	26	40	45	50	51	52	53
$d_3(x)$	0	1	2	3	4	5	6	7	8
$f_2(x)$	0	5	26	40	60	70	86	100	110
$d_2(x)$	0	1	0	0	4	5	4	4	5

第三步：若有三个资源使用者，则设

$f_1(x)$——为 x 单位的资源分配到三个使用者时的最大获益.

z——为第一个使用者获得使用的资源.

$x-z$——为第二、三个使用者获得使用的资源.

则有：$f_1(x)=\max\{g_1(z)+f_2(x-z)\}$

其中 $z=0,1,2,\cdots,x$

$d_1(x)=z \quad (x=0,1,\cdots,8)$ 且满足 $f_1(x)$.

当 x 取不同值时，利用表 4-2 和表 4-4，可求得不同的 $f_1(x)$ 及 $d_1(x)$.

例如 $x=7$ 时，有

$$f_1(7)=\max_{z=0,1,\cdots 7}\{g_1(z)+f_2(x-z)\}$$

$$=\max\begin{Bmatrix} g_1(0)+f_2(7) \\ g_1(1)+f_2(6) \\ g_1(2)+f_2(5) \\ g_1(3)+f_2(4) \\ g_1(4)+f_2(3) \\ g_1(5)+f_2(2) \\ g_1(6)+f_2(1) \\ g_1(7)+f_2(0) \end{Bmatrix}$$

$$=\max\begin{Bmatrix} 0+100 \\ 5+86 \\ 15+70 \\ 40+60 \\ 80+40 \\ 90+26 \\ 95+5 \\ 98+0 \end{Bmatrix}$$

$$=120$$

$$d_1(7)=4$$

这里 $f_1(7)$是将总资源中的 7 个单分配给三个使用者的最大收益，其中第一位使用者占有 4 个单位的资源，而剩下的 3 个单位的资源由表 4-4 知，应由第二个使用者全部占有.

即 $d_1(7)=4$ 时，$d_2(7-4)=3$，$d_3(0-0)=0$，此时 $f_1(7)=120$，由表 4-2 可知：

$f_1(7)=g_1(4)+g_2(3)+g_3(0)=80+40+0=120$ 经过计算，我们可得 0～8 个单位的资源分配到各个使用者时最佳分配方案，列表 4-5.

表 4-5

x	0	1	2	3	4	5	6	7	8
$f_3(x)$	0	4	26	40	45	50	51	52	53
$d_3(x)$	0	1	2	3	4	5	6	7	8
$f_2(x)$	0	5	26	40	60	70	86	100	110
$d_2(x)$	0	1	0	0	4	5	4	4	5
$f_1(x)$	0	5	26	40	80	90	106	120	140
$d_1(x)$	0	0	0	0	4	5	4	4	4

由表 4-5 可以得到允许分配方案中的任一最优分配，例如：总资源为 8 个单位时，所获得的最大收益为 $f_1(8)=140$，而此时第一个使用资源仅得到 4 个单位的资源即 $d_1(8)=4$，第二个使用者获得 4 个单位的资源即 $d_2(8-4)=4$. 第三个使用者没有获得资源使用权 $d_3(4-4)=0$，因此最大收益应由：

$$f_1(8)=g_1(4)+g_2(4)+g_3(0)=80+60+0=140$$

决定.

将资源分配问题概括为动态规划可以表示为：

当有 K 个资源，分配到 N 个使用者时，其规划方程

$$\begin{cases} f_n(x)=g_n(x) \\ d_n(x)=x \end{cases} \quad (x=0,1,\cdots k)$$

其子过程是

$$f_i(x)=\max_{y=0,1,\cdots,x}\{g_i(y)+f_{i+1}(x-y)\}$$

$$d_i(x)=\text{产生 } f_i(x) \text{ 的 } y \text{ 值} \quad (i=1,2,\cdots,n-1;x=0,1,\cdots,k)$$

习题四

1. 加工某食品 100 斤，需 A、B、C 三种原料，且含 A 不少于 40 斤，含 B 不多于 20 斤，含 C 不多于 30 斤，而 A、B、C 每斤成本分别为 0.5 元，0.3 元，0.4 元. 问：

(1)确定 A、B、C 成分数量是多少时成本最低。

(2)每生产 100 斤食品的最低成本。

2. 用图解法和单纯形法求解下列线性规划问题。

(1)
$$\max \quad Z=2x_2+3x_2$$

$$\text{约束} \quad -x_1+x_2\leqslant 1$$

$$-2x_1+10x_2\leqslant 20$$

(2) $\max \quad Z = 2x_1 + 5x_2$

$$2x_2 \leqslant 12$$

$$3x_1 + 2x_2 \leqslant 18$$

$$x_1, x_2 \geqslant 0$$

3. 设某工厂生产 A、B、C 三种零件，每种零件的材料单耗、所需加工工时，资源月限量及每件零件的产值如表(4-6)。根据实际情况，A、B、C 三种零件的月生产产量分别控制在：200～250，250～280，100～120 件。如何安排生产这三种零件的数量，使工厂的月产值最大？

零 件	材料消耗(kg)	加工工时(小时)	单位零件产值(元)
A	1.0	2.0	10
B	1.5	1.2	14
C	4.0	1.0	12
资源月限量	2000	1000	

4. 设某广告公司对一产品所做广告费计划为 120 000 元，并选定两种广告方式，第一种方式每次需 2 000 元，第二种方式每次需 5 000 元。由经验可知，这两种广告方式分别至少实现 20 次和 10 次，同时总次数不超过 50 次时，才能收到预期的效果。问：怎样设计各种广告的次数组合，才能使总的广告费不超过预算？

5. 用二阶段法求解线性规划问题。

$\max \quad Z = 40x_1 + 60x_2$

约束 $2x_1 + 4x_2 \leqslant 180$

$$3x_1 + 2x_2 \leqslant 150$$

$$x_1 + x_2 = 57$$

$$x_2 \geqslant 22$$

$$x_1, x_2 \geqslant 0$$

6. 设某矿有 5 个出煤井口，位置已知，矿区平面示意图如图 4-3。在图示阴影地区选定一煤处理厂址，使之到各井口的距离总和最小。

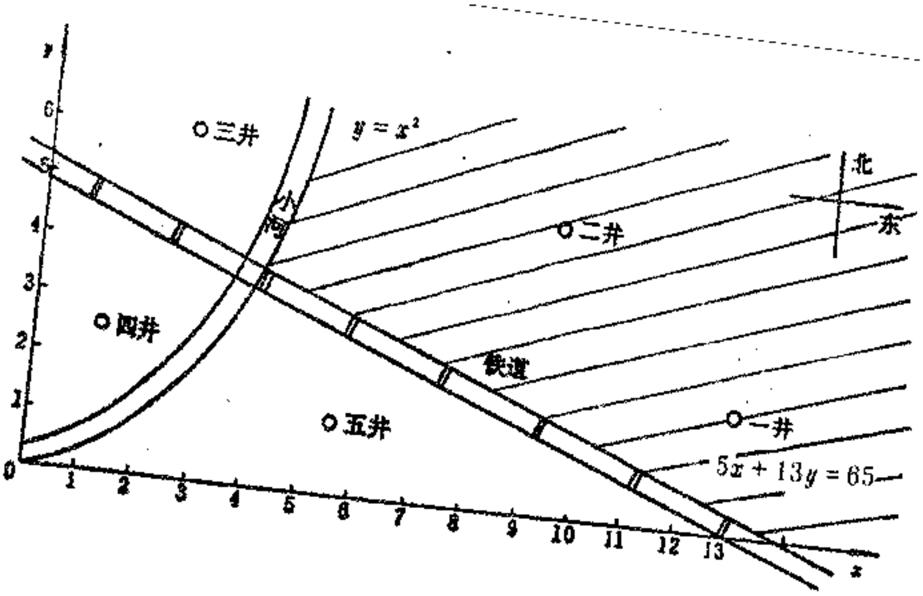


图 4-3

7. 甲乙二公司通过做广告来竞争产品的销售量，广告费分别是 x,y，两公司产品在市场上销售所占的比例是其广告费在总广告费中所占比例的函数。公司收入与销售量成正比，从收入中扣除广告费是公司的利润。问甲、乙两公司怎样确定广告费并使其利润最大？

8. 某人外出旅游，需将五件物品装入包裹，但包裹总重量限制在 13 公斤以内。物品的重量及价值如表 4-7，问如何选装这些物品，使包裹的价值最大？

物　品	重量(公斤)	价值(元)
1	7	9
2	5	4
3	4	3
4	3	2
5	1	0.5

9. 某煤矿月生产能力 5 万吨，由于运力的限制，超产部分需要地面存放，且不超 4 个月。如果生产费用为 3 元/吨，生产附加费 6 万元/月。地面存贮费 1.5 元/吨。问如何安排生产，使总费用最少？

10. 某工厂有机器 250 台，计划使用 4 年。若有两项生产任务需要完成，且机器用于第一项生产任务每台机器可获利 850 元，损坏率为 35%；用于第二项生产任务每台机器可获利 650 元，损坏率 20%。问在四年内如何使用这些机器并获得最好的效益？利润是多少？

11. 用动态规划的方法求解线性规划问题

$$\max \quad Z = 4x_1^2 - x_2^2 + 2x_3^2 + 12$$

$$约束 \quad 3x_1 + 2x_2 + x_3 = 9$$

$$x_i \geq 0, i = 1,2,3$$

12. 计算从 A 到 B、C、D 的最短路线。已知各段路线的长度如图 4-4 所示。

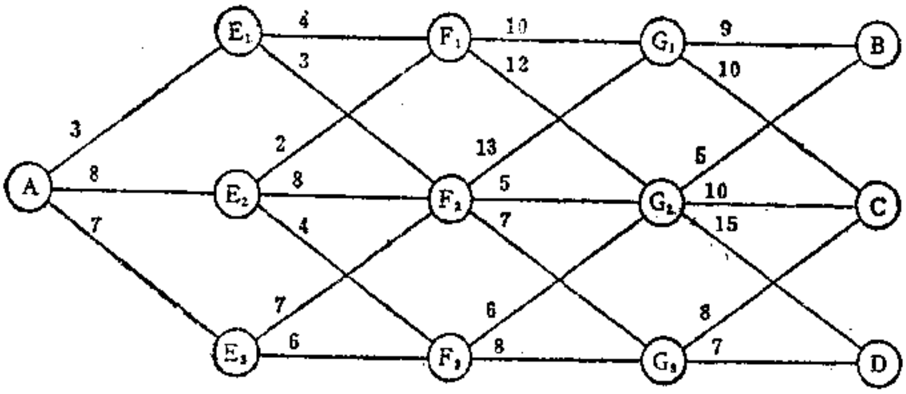


图 4-4

10

第五章　数理统计模型

数理统计是以概率论为理论基础，根据试验或观察得到的数据，对研究的对象的客观规律作出各种合理的估计、分析、检验和推断．数据统计包含的内容相当广泛，诸如统计推断、多元分析、试验设计、时间序列分析等，其具体方法甚多，应用相当广泛．由于篇幅所限，本章主要介绍随机决策模型，多元线性回归分析、广义线性回归分析、方差分析．大家知道，数理统计模型的分析计算量相当大，既使用计算机编程计算需要的时间也相当多，因此我们提倡用已有的计算软件如：SAS、Mathematilc、MathCAI 等来完成这些计算．

第一节　随机决策模型

所谓决策，就是从多个备选方案中，选择一个最优的或满意的方案付诸实施．决策问题依据在决策过程中是否使用数学模型，可分为数理性决策与非数理性决策．非数理性决策，主要依靠传统与习惯进行，或依靠经验与直觉进行，一般无法使用数学模型；数理性决策问题可分为确定性情况与非确定性情况，确定性决策中的条件与状态是比较清楚的，变动因素可以用确定性变量来描述，可建立确定性模型来求解，如线性规划、动态规划等．非确定性决策也称为随机决策，一般可分为风险型决策与定全不确定型决策．

风险型决策中存在着非确定性因素，决策活动要考虑几种不同的外部条件，也称为“自然状态”，例如，市场需求、气候变化、政治形势等．哪一种自然状态将出现是不确定的，但每一种状态出现的可能性，也称为“主观概率”，有一个估计值；完全不确定型决策不仅存在着几种自然状态（与风验型决策相同），而且各种自然状态出现的可能性即主观概率也无法确定（与风险型决策不同）．本节重点讨论随机决策模型中的风险型决策与完全不定型决策．

一、决策模型

下面请看两个实例

例 1　某厂要确定下一计划期内产品的生产批量，根据以前经验并通过市场调查和

表 5-1　　单位：千元

效益值 \ 自然状态 \ 概率 \ 行动方案	产品销路		
自然状态	N_1(好)	N_2(一般)	N_3(差)
概率	$P_1=0.4$	$P_2=0.5$	$P_3=0.1$
A_1(大批量生产)	30	15	4
A_2(中批量生产)	20	20	12
A_3(小批量生产)	10	15	14

预测，已知产品销路好、一般、差三种情况的可能性分别为 0.4、0.5 和 0.1，产品采用大、

中、小批量生产，可能获得的效益值也可以相应地计算出来，详见表 5-1，现要通过决策分析，确定合理批量，使企业获得效益最大？

例 2 某公司为扩大市场，要举办一个产品展销会，会址打算选择甲、乙、丙三地，获利情况除了与会址有关系外，还与天气有关，天气分为晴、阴、多雨三种，据气象预报，估计三种天气情况可能发生的概率为 0.2，0.5，0.3，其收益情况见表 5-2，现要通过分析，确定会址，使收益最大。

表 5-2 单位：万元

收益值 \ 自然状态 / 概率 \ 选址方案	天气情况		
	N_1(晴)	N_2(阴)	N_3(多雨)
	$P_1=0.20$	$P_2=0.50$	$P_3=0.30$
A_1(甲地)	4	6	1
A_2(乙地)	5	4	1.5
A_3(丙地)	6	2	1.2

在决策问题中，把面临的几种自然情况称为自然状态或客观条件，简称为状态或条件，如例 1、例 2 中的 N_1, N_2, N_3，这些是不可控因素，把 A_1, A_2, A_3 称为行动方案或策略，这些是可控因素，至于选择哪个方案由决策者决定。表 5-1 中右下方的数字 30，15，4，20，20，12，10，15，14，表 5-2 中右下方的数字 4，6，1，5，4，1.5，6，2，1.2 称为益损值，根据这些数字的含义不同，有时也称为效益值或风险值。由它们构成的矩阵

$$M_1=\begin{pmatrix}30 & 15 & 4\\ 20 & 20 & 12\\ 10 & 15 & 14\end{pmatrix} \qquad M_2=\begin{pmatrix}4 & 6 & 1\\ 5 & 4 & 1.5\\ 6 & 2 & 1.5\end{pmatrix}$$

称为决策的益损矩阵或风险矩阵。表 5-1 和表 5-2 中的 P_1, P_2, P_3 是各状态出现的概率。

一般地，如果决策问题的可控因素，即行动方案用 $A_i(i=1\sim m)$ 表示，状态用 $N_j(j=1\sim n)$ 表示，在 N_j 状态下采用 A_i 行动方案的益损值用 a_{ij} 表示，N_j 状态下的概率用 $p_j(j=1\sim n)$ 表示，可以得到决策矩阵（或称益损矩阵）的一般结构，如表 5-3 所示。

表 5-3

益损值 \ 状态 / 概率 \ 方案	N_1	N_2	…	N_j	…	N_n
	P_1	P_2	…	P_j	…	P_n
A_1	a_{11}	a_{12}	…	a_{1j}	…	a_{1n}
…	…	…	…	…	…	…
A_i	a_{i1}	a_{i2}	…	a_{ij}	…	a_{in}
…	…	…	…	…	…	…
A_m	a_{m1}	a_{m2}	…	a_{mj}	…	a_{mn}

易见，当 $n=1$ 时，决策问题就是确定型的。下面分别讨论风险型和不确定型的决策问题。

二、风险型决策

当表 5-3 中的 $n>1$，且各种自然状态下出现的概率 $P_i(i=1\sim n)$ 已知时的决策问题就是风险型决策问题. 如例 1、例 2 的决策问题，均属于风险型决策问题. 求解这类问题可以运用最大可能性准则和期望值准则，也可采用决策树法.

1. 最大可能性准则

由概率论知识知道，一个事件的概率就是该事件在一次试验中发生的可能性大小，概率越大，发生的可能性就越大. 基于这种思想，在风险决策中，先选择发生概率最大的自然状态列(不顾及其它自然状态)，再从该列选择收益最大(或损失最小)的方案为最优方案. 这个准则的实质是将风险型决策问题转化为确定型决策问题的一种决策方法.

例 3 应用最大可能准则求解例 1.

解 因为自然状态 N_2 出现的概率 $P_2=0.5$ 最大，因此就在这种自然状态下进行决策，这时表 5-1 变成了表 5-4，即变成了确定型决策问题.

表 5-4

方案 \ 益损值 \ 状态	产品销路 N_2(一般)
A_1(大批量生产)	15
A_2(中批量生产)	20
A_3(小批量生产)	15

通过比较表 5-4 中最右一列数值可知，采取 A_2 行动方案获利最大，所以选取 A_2(即中批量生产)是最优方案.

这个准则用起来比较方便. 但要注意，在一组自然状态中，其中某一个状态出现的概率比其它状态出现的概率特别大，而它们相应的益损值差别不很大，这种准则的效果是较好的. 如果在一组自然状态中，它们发生的概率都很小，而且互相很接近，再要采用这种决策准则，效果是不好的，有时可能造成较大的失误.

2. 期望值准则

该准则是，在决策矩阵(表 5-3)中计算每一方案的益损期望值，从中选取收益最大值者(或损失最小值者)为最优方案. 第 i 个方案的期望值计算公式为：

$$E(A_i)=\sum_{j=1}^{n}a_{ij}\cdot P_j,\qquad i=1\sim m.$$

例 4 应用期望值准则求解例 1.

$E(A_1)=30\times0.4+15\times0.5+4\times0.1=19.9$

$E(A_2)=20\times0.4+20\times0.5+12\times0.1=19.2$

$E(A_3)=10\times0.4+15\times0.5+14\times0.1=12.9$

通过比较可知 $E(A_1)=19.9$ 最大，所以采用行动方案 A_1，也就是采取大批量生产，可能获得的效益最大.

3. 决策树法

决策树法就是把某个决策问题未来发展情况的可能性和可能结果所做的预测用树状图画出来(形象地称为决策树)，然后根据期望值准则进行决策的一种方法.

(1) 单级决策问题

下面以例 2 来说明其决策步骤.

由前面表 5-2 数据可作出决策树如图 5-1 所示.

图 5-1 中符号说明：

□——表示决策点，从它引出的分枝叫方案分枝，分枝数反映可能的行动方案数；

○——表示方案节点，其上方的数字表示该方案的益损期望值，从它引出的分支叫概率分支，每条分支的上面写明了自然状态及其出现的概率，分枝数反映可能的自然状态数；

△——表示末稍节点，它旁边的数字是每一方案在相应状态下的益损值.

将各方案节点上的期望值加以比较，选取最大的益损期望值 4.1，写在决策点的上方，说明选定了方案 A_1，方案分枝中打有"$+\!\!+$"记号的表示该方案被删除掉或称剪枝方案.

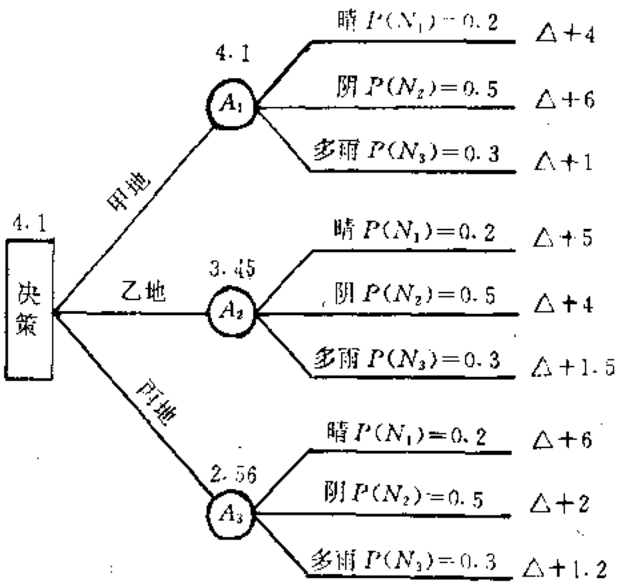


图 5-1 决策树

注意 决策问题的目标如果是效益（如利润、投资、回报等）应取期望值的最大值. 如果决策目标是费用的支出或损失，则应取期望值的最小值.

(2) 多级决策问题

例 5 为生产某种产品，设计了两个基建方案，一是建大厂，二是建小厂，大厂需要投资 300 万元，小厂需要投资 160 万元，两者的使用期都是 10 年. 估计在此期间，产品销路好的可能性是 0.7，两个方案的年度益损值见表 5-5.

表 5-5 单位：万元

自然状态	概 率	建大厂	建小厂
销路好	0.7	100	40
销路差	0.3	−20	10

(a) 画决策树（图 5-2）

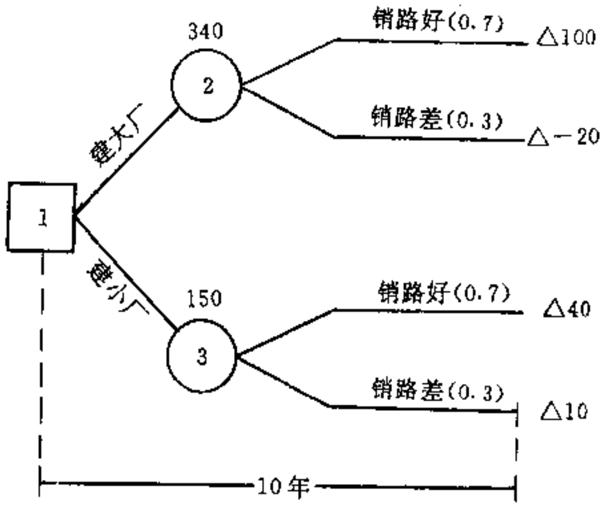


图 5-2　决策树

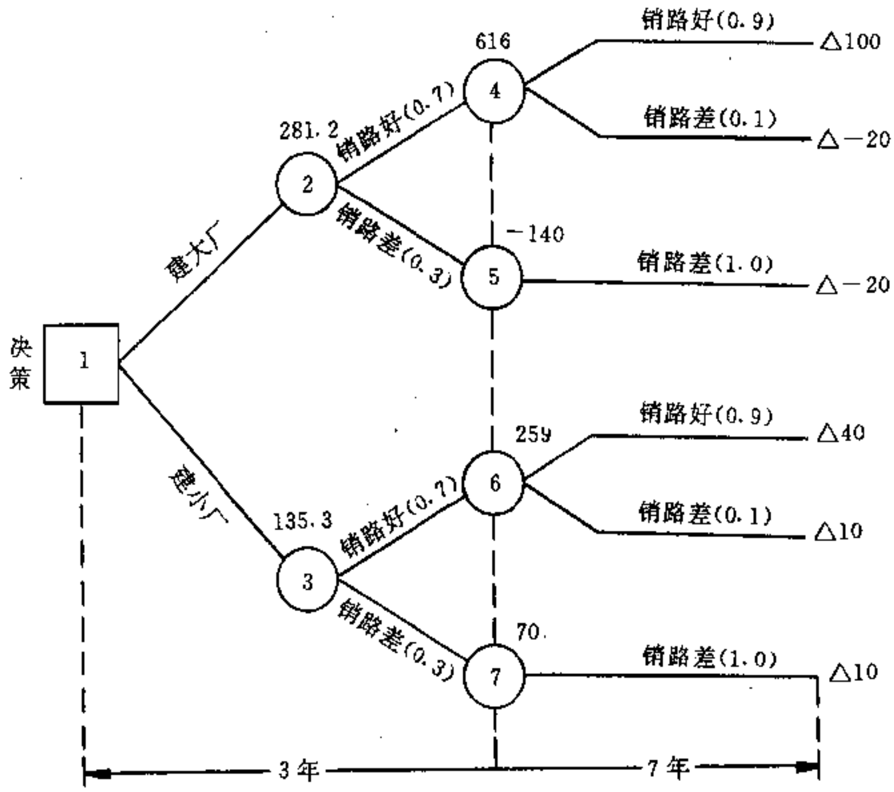


图 5-3　决策树

(b) 计算各点的益损期望值：

点 2：[0.7×100+0.3×(−20)]×10(年)−300(大厂投资)=340 万元

点 3：[0.7×40+0.3×10]×10(年)−160(小厂投资)=150 万元

由此可见,建大厂的方案是合理的.

现在考虑一种情况:

假定对例 5 分为前三年和后七年两期考虑.根据市场预测,前三年销路好的概率为 0.7,而如果前三年的销路好,则后七年销路好的概率为 0.9,如果前三年的销路差,则后七年的销路肯定差,在这种情况下,建大厂和建小厂那个方案好?

(a) 画出决策树如下(图 5-3)

(b) 计算各点的益损期望值

点 4:[0.9×100+0.1×(−20)]×7(年)=616

点 5:1.0×(−20)×7(年)=−140

点 2:0.7×100×3(年)+0.7×616+0.3×(−20)×3(年)+0.3×(−140)

−300(大厂投资)=281.2

点 6:[0.9×40+0.1×10]×7(年)=259

点 7:1.0×10×7(年)=70

点 3:0.7×40×3(年)+0.7×259+0.3×10×3(年)+0.3×70

−160(小厂投资)=135.3

通过比较,建大厂仍然是合理方案.

思考题:若先建设小工厂,如销路好,则三年以后考虑扩建.扩建投资需要 140 万元,扩建后可使用七年,每年的益损值与大厂相同.这个方案与建大厂方案比较,优劣如何?

三、完全不确定型决策

当风险决策问题的自然状态发生的概率既不知道,也无法预先估计或利用历史资料得到时的决策问题就称为完全不确定型决策问题.

例 6 如果在例 1 给定的条件中,只知道产品销路有三种可能,但三种销路究竟各有多大的可能性却不知道,这时的给定条件可以用表 5-6 表示.试确定该企业采用哪种批量生产,使企业获得的效益最大?

为了对例 6 进行决策,下面介绍几种不确定型决策准则.

1. 乐观准则

也称最大准则或大中取大准则(Maxi max Criterion),这种准则的思想基础是对客观情况总是抱乐观态度.具体方法是:首先从每一个行动方案中选取一个最大的效益值,然后再从这些最大效益值中找出最大的,它所对应的方案即为最优方案.

(注:若是损失矩阵,则应取最小最小准则)

表 5-6

自然状态 / 效益值 / 行动方案	产品销路		
	N_1(好)	N_2(一般)	N_3(差)
大批量生产 A_1	30	15	4
中批量生产 A_2	20	20	12
小批量生产 A_3	10	15	14

在例 6 中，由表 5-6 可知各行动方案的最大效益值分别为：

$\max\limits_{j}\{a_{1j}\}=\max\limits_{A_1}\{30,15,4\}=30$

$\max\limits_{j}\{a_{2j}\}=\max\limits_{A_2}\{20,20,12\}=20$

$\max\limits_{j}\{a_{3j}\}=\max\limits_{A_3}\{10,15,14\}=15$

各行动方案最大效益中的最大值为

$$\max_i\{\max_j \quad a_{ij}\} = \max\{30,20,15\} = 30$$

其最大值 30 对应的行动方案为 A_1，因此用乐观法的决策结果是采用大批量生产方案.

2. 悲观准则

也称最大最小准则或小中最大准则(Maxi min Criterion)，这种准则的思路是对客观情况总是抱悲观态度，万事总觉得不如意. 具体方法是：首先从每一个行动方案中选取一个最小的效益值，然后再从这种最小效益值中找出最大的，它所对应的方案即为最优方案.

(注：若是损失矩阵，则应取最小最大准则)

在例 6 中，由表 5-6 可知各行动方案的最小效益值分别为：

$\min\limits_{j}\{a_{1j}\}=\min\limits_{A_1}\{30,15,4\}=4$

$\min\limits_{j}\{a_{2j}\}=\min\limits_{A_2}\{20,20,12\}=12$

$\min\limits_{j}\{a_{3j}\}=\min\limits_{A_3}\{10,15,14\}=10$

各行动方案最小效益中的最大值为

$$\max_i\{\min_j \quad a_{ij}\} = \max\{4,12,10\} = 12$$

其最大值 12 对应的行动方案为 A_2，因此用悲观法的决策结果是采用中批量生产方案.

3. 等可能性准则

也称理由不充分准则，这种准则的思想是，既然不能断定那种自然状态出现的可能性(概率)的大小，就认为各自然状态出现的可能性相同，即 $P(N_j)=\frac{1}{n}(j=1\sim n)$，然后按风险型决策的方法进行决策.

在例 6 中，取 $P(N_j)=\frac{1}{3}$，$j=1,2,3$，计算各行动方案的期望值为：

$E(A_1)=30\times\frac{1}{3}+15\times\frac{1}{3}+4\times\frac{1}{3}=\frac{49}{3}$

$E(A_2)=20\times\frac{1}{3}+20\times\frac{1}{3}+12\times\frac{1}{3}=\frac{52}{3}$

$E(A_3)=10\times\frac{1}{3}+15\times\frac{1}{3}+14\times\frac{1}{3}=\frac{39}{3}$

$\max\{E(A_1),E(A_2),E(A_3)\}=\max\{\frac{49}{3},\frac{52}{3},\frac{39}{3}\}=\frac{52}{3}$

易见，用等可能性准则的决策结果是采用中批量生产方案.

4. 折中准则

也称乐观系数准则，这个准则的特点是，对客观条件的估计既不那么乐观，但也不悲观，而是根据历史数据的分析和经验判断确定一个“乐观系数”$\alpha\in[0,1]$，然后对每一个行动方案作如下计算：

$$折中效益值 = \alpha \times 最大效益值 + (1-\alpha) \times 最小效益值$$
$$= \alpha \max_j\{a_{ij}\} + (1-\alpha)\min_j\{a_{ij}\}$$

然后在折中效益值中选择最大值所对应的行动方案为最优方案.

注：如果考虑的是损失值，则

$$折中损失值 = \alpha \min_j\{a_{ij}\} + (1-\alpha)\max_j\{a_{ij}\}$$

易见 $\alpha=1$ 是乐观准则，$\alpha=0$ 是悲观准则.

今取 $\alpha=0.7$，以例 6 来说明.

折中效益值 $E_1=0.7\times\max\{30,15,4\}+0.3\times\min\{30,15,4\}=22.2$

折中效益值 $E_2=0.7\times\max\{20,20,12\}+0.3\times\min\{20,20,12\}=17.6$

折中效益值 $E_3=0.7\times\max\{10,15,14\}+0.3\times\min\{10,15,14\}=13.5$

取折中效益值中的最大值

$$\max\{22.2,17.6,13.5\} = 22.2$$

易见，用折中准则的决策结果是采用大批量生产方案.

5. 后悔值准则

决策者制定决策之后，若情况未能符合理想，必将有后悔的感觉，或者感到遗憾，所以也称遗憾准则. 这种决策准则的出发点是将每种自然状态的最高值(指收益矩阵，损失矩阵取最低值)订为该状态的理想目标，并将该状态中的其他值与最高值相减所得之差，称为未达到理想之后悔值. 后悔准则就是在决策时尽量减少将来的后悔. 具体做法是，首先将每一种自然状态下的最大收益减去其他收益，即由收益矩阵求得后悔矩阵；再从每一个行动方案中找出最大后悔值；最后从这些最大后悔值中找出最小的，它所对应的方案既为最优方案.

仍以例 6 来说明.

首先从收益矩阵计算出后悔矩阵，见表 5-7.

表 5-7

收益矩阵 / 行动方案	自然状态		
	N_1	N_2	N_3
A_1(大批量生产)	30*	15	4
A_2(中批量生产)	20	20*	12
A_3(小批量生产)	10	15	14*
后悔矩阵			
A_1(大批量生产)	0	5	⑩
A_2(中批量生产)	⑩	0	2
A_3(小批量生产)	⑳	5	0

从每一个行动方案中找出最大后悔值，即从后悔矩阵中每一行选一个后悔值最大的数，第 1 行(大批量生产)为 10，第 2 行(中批量生产)为 10，第 3 行(小批量生产)为 20，再

从这些最大后悔值中选最小值

$$\min\{10,10,20\} = 10$$

所以，按后悔值准则的决策结果是采用大批量(或中批量)生产即为最优方案.

6. 意愿准则

当备选方案数目很多时，可以用意愿准则进行初选，淘汰若干方案. 具体说就是：决策者根据自己拥有的资源条件和经营目标，对方案的收益值规定一个最低限(或对方案的投资规定一个最高限)，凡超过规定限额的方案均予淘汰，然后在剩下的方案中再进行决策.

比如例 6，若规定收益值$\geqslant 10$，则淘汰方案 A_1，因为 A_1 在 N_3 时的收益为 $4<10$；若规定收益值$\geqslant 12$，则唯一可供进一步选择的方案只有 A_2.

对于不确定型的决策问题，采用不同的决策准则所得到的结果并非完全一致，由于没有一个统一的评比标准，所以难以判别那个准则好，那个准则不好，究竟在实际问题中采用那个准则，依决策者对各种自然状态的看法而定. 当然，为了改进不确定型决策，必须设法获得各种自然状态发生的可能性，以便使决策的结果可能更合理一些.

第二节　多元线性回归分析

在现实生活中，变量与变量之间经常存在一定的关系. 一般来说，变量之间的关系可以分为两大类，一类是确定性的关系，这种关系通常用函数来表示. 例如，已知圆的半径 r，那么圆的面积 S 与半径 r 的关系就可用函数关系：$S=\pi r^2$ 来表示，这时如果取定了 r 的值，S 的值就会完全确定了. 另一类是非确定性关系，例如，人的体重与身高之间的关系就是非确定性关系. 一般来说，身高越高，体重越大，但是身高相同的人体重往往是不相同的. 再如，钢材的强度与钢材中含某种元素的含量，纤维的拉伸倍数与强度，降雨量、气温、施肥量与农作物的产量等均属于这种关系. 变量之间的这种非确定性关系通常称为相关关系.

回归分析就是数理统计中研究相关关系的一种数学方法，它就是通过大量的试验或观测，发现变量之间关系的统计规律.

为了研究方便，我们考虑一个变量受其它变量影响时，仍把这变量称为因变量，记为 Y，其它变量称为自变量，记为 X，这时相关关系可记作

$$Y = f(x) + \varepsilon \tag{5.1}$$

其中 $f(x)$为当 $X=x$ 时，因变量 Y 的均值，即

$$f(x) = E(Y|X = x)$$

称 $f(x)$为 Y 对 X 的回归函数，ε 为 Y 与 $f(x)$的偏差，它是一个随机变量，并假定 $E(\varepsilon)=0$.

"回归"一词始于英国统计学家盖尔登(Galton)和皮尔逊(Pearson). 他们研究儿子的身高 Y 与其父母亲的平均身高 X 之间的关系. 通常双亲的平均身高较矮时，其子的身高也较矮；双亲的平均身高较高时，其子的身高也较高，但是他们发现，如果双亲属于高个类(高于 1000 多对夫妇的平均身高)时，其子比他父亲更高的概率会较小；相反地，如果双亲属于低个类时，其子比他父亲更矮的概率也较小. 所以，平均身高偏高或偏矮的双亲，共子的身高都有一种"向中心回归"的现象，基于这个事实，盖尔登把所求描述儿子与双亲身高

关系的直线称为回归直线.

回归函数可以是一元函数,也可以是多元函数,即

$$Y=f(x_1,x_2,\cdots,x_m)+\varepsilon \tag{5.2}$$

其中 $f(x_1,x_2,\cdots,x_m)=E(Y|X_1=x_1,X_2=x_2,\cdots,X_m=x_m)$ 为 m 元回归函数,统称为多元回归函数.

若回归函数 $f(x_1,x_2,\cdots,x_m)$ 中的 $m=1$,且是一元线性函数,则称为是一元线性回归;$m>1$ 且是多元线性函数,则称为是多元线性回归;若回归函数 $f(x_1,x_2,\cdots,x_m)$ 是非线性函数,则称为是非线性回归.对非线性回归,经常采用线性化的方法来处理.所以,目前研究最多的是线性回归,且假定 $X_1,X_2,\cdots,X_m$ 和 Y 均服从正态分布.回归分析的任务就是要求出满足式(5.2)的回归函数 $f(x_1,x_2,\cdots,x_m)$,从而对所研究的相关关系作出所需的预测和控制.

一、多元线性回归的数学模型

多元回归模型的应用是相当广泛的,例如,某种商品的销售量可能受收入水平、风俗习惯、产品质量、价格、宣传广告等多种因素的影响;某种产品的质量可能受生产该产品时的温度、湿度、压力、原材料的质量和有害成份的含量等影响;工人的劳动生产率可能受学历、智力水平、情绪的稳定性和才能等因素的影响;某城市的用水量可能与该城市的人口数及工业总产值有关.诸如此类的关系,可以通过多元回归分析模型进行研究.

下面先看一个实例.

例 1 某造纸公司会计部门根据长期积累的资料发现,公司每月管理费用 y 与工人的劳动日数 x_1 及机器的日平均开工台数 x_2 有关,现将某年的 1~10 月份的数据整理于表 5-8,试研究管理费用与工人的劳动日数和机器的日平均开工台数的关系.

表 5-8

月 份	管理费用 y	劳动日数 x_1	机器平均开工台数 x_2
1	29	45	16
2	24	42	14
3	27	44	15
4	25	45	13
5	26	43	13
6	28	46	14
7	30	44	16
8	28	45	16
9	28	44	15
10	27	43	15

一般来说,劳动日数越多,管理费用越高;机器平均开工台数越多,管理费用越高,于是可以建立线性回归模型

$$Y = b_0 + b_1x_1 + b_2x_2 + \varepsilon \tag{5.3}$$

其中 $E(\varepsilon)=0, D(\varepsilon)=\sigma^2, b_0, b_1, b_2$ 和 σ^2 是未知参数，为了估计这些参数，将表 5-8 的值代入模型(5.3)，得线性模型

$$\begin{cases} y_i = b_0 + b_1x_{i1} + b_2x_{i2} + \varepsilon_i \\ E(\varepsilon_i) = 0, \quad \mathrm{Cov}(\varepsilon_i, \varepsilon_j) = \delta_{ij}\sigma^2 \quad (i, j = 1 \sim 10) \end{cases} \tag{5.4}$$

一般地，多元线性回归模型可表示为

$$Y = b_0 + b_1x_1 + b_2x_2 + \cdots + b_mx_m + \varepsilon \tag{5.5}$$

其中，$x_1, x_2, \cdots, x_m$ 是自变量，b_0 为常数，$b_1, b_2, \cdots, b_m$ 为回归系数，$b_0, b_1, b_2, \cdots, b_m$ 皆为未知，统称 $b_0, b_1, b_2, \cdots, b_m$ 为回归参数，一旦回归参数确定，则多元线性回归模型就完全确定. 一般假定随机误差 $\varepsilon \sim N(0, \sigma^2)$.

为了得到回归参数的估计值，就要对变量进行观测，假设对变量的 $n(n>m)$ 次独立观测数据为

$$(y_i, x_{i1}, x_{i2}, \cdots, x_{im}), i = 1 \sim n,$$

则这些观测数据应满足式(5.5)，即有

$$\begin{cases} y_1 = b_0 + b_1x_{11} + \cdots + b_mx_{1m} + \varepsilon_1 \\ y_2 = b_0 + b_1x_{21} + \cdots + b_mx_{2m} + \varepsilon_2 \\ \cdots \\ y_n = b_0 + b_1x_{n1} + \cdots + b_mx_{nm} + \varepsilon_n \end{cases} \tag{5.6}$$

其中 $E(\varepsilon_i)=0, \mathrm{Cov}(\varepsilon_i, \varepsilon_j)=\delta_{ij}\sigma^2 \quad i, j=1\sim n.$

若记 $Y=(y_1, y_2, \cdots, y_n)^T, \beta=(b_0, b_1, \cdots, b_m)^T, \varepsilon=(\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n)^T$,

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1m} \\ 1 & x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nm} \end{pmatrix}_{n\times(m+1)}$$

则多元线性回归的数学模型式(5.6)可以写成矩阵形式

$$Y = X\beta + \varepsilon \tag{5.7}$$

其中 $E(\varepsilon)=0, \mathrm{Var}(\varepsilon)=\sigma^2 I_n$.

二、参数的最小二乘估计

为了获得参数 β 的估计，我们采用最小二乘法，即选择 β，使

$$\begin{aligned} Q(\beta) &= \sum_{i=1}^{n} \varepsilon_i^2 \\ &= \varepsilon^T \varepsilon \\ &= (Y - X\beta)^T (Y - X\beta) \end{aligned} \tag{5.8}$$

达到最小.

将 $Q(\beta)$ 对 β 求导数并令其为零，得

$$\frac{\partial Q}{\partial \beta} = -2X^T(Y - X\beta) = 0$$

即
$$X^TX\beta = X^TY$$

记　$L=X^TX$,则
$$L\beta = X^TY \tag{5.9}$$

方程(5.9)称为正规方程.其中 X 为 $n\times(m+1)$ 阶矩阵,一般假定 $\text{rank}(X)=m+1$,由线性代数理论可知,$L=X^TX$ 为满秩矩阵,它的秩 $\text{rank}(L)=m+1$,则正规方程(5.9)有唯一解,记作
$$\hat{\beta}= L^{-1}X^TY \tag{5.10}$$

我们来证明(5.10)式中的 $\hat{\beta}$ 为参数向量 β 的最小二乘法估计量,现用矩阵形式来叙述其证明步骤.

从式(5.8)知,对任意的 β
$$Q = (Y - X\beta)^T(Y - X\beta)$$

则有
$$\begin{aligned}(Y - X\beta)^T(Y - X\beta) &= [(Y - X\hat{\beta}) + X(\hat{\beta}-\beta)]^T[(Y - X\hat{\beta}) + X(\hat{\beta}-\beta)] \\ &= (Y - X\hat{\beta})^T(Y - X\hat{\beta}) + (\hat{\beta}-\beta)^TX^TX(\hat{\beta}-\beta) \\ &\quad + (Y - X\hat{\beta})^TX(\hat{\beta}-\beta) + (\hat{\beta}-\beta)^TX^T(Y - X\hat{\beta}) \geqslant (Y - X\hat{\beta})^T(Y - X\hat{\beta})\end{aligned}$$

上述证明过程中应用了如下结果:
$$(\hat{\beta}-\beta)^TX^TX(\hat{\beta}-\beta) = [X(\hat{\beta}-\beta)]^T[X(\hat{\beta}-\beta)] \geqslant 0$$

$$\begin{aligned}(Y - X\hat{\beta})^TX(\hat{\beta}-\beta) &= (Y^TX - \hat{\beta}^TX^TX)(\hat{\beta}-\beta) \\ &= (Y^TX - Y^TX)(\hat{\beta}-\beta) = 0\end{aligned}$$

至此,在 $|L|\neq 0$ 时,证明了式(5.10)中的 $\hat{\beta}$ 是 β 的最小二乘法估计量.

在实际工作中,常称
$$\hat{y}= \hat{b}_0 + \hat{b}_1x_1 + \cdots + \hat{b}_mx_m \tag{5.11}$$
为经验线性回归方程.

三、最小二乘法估计量的性质

首先我们在假定 $E(\varepsilon)=0$, $\text{Var}(\varepsilon)=\sigma^2I_n$ 的条件下,探讨一下由式(5.10)确定 β 的最小二乘法估计量 $\hat{\beta}$ 的性质.

1. $\hat{\beta}$ 是 β 的线性无偏估计量.

证:由于 $\hat{\beta}=L^{-1}X^TY$,每一个 $\hat{b}_i$ 都是 $y_1,\cdots,y_n$ 的线性组合,因而 $\hat{b}_i$ 是 b_i 的线性估计量,此时称 $\hat{\beta}$ 是 β 的线性估计量.
$$\begin{aligned}E(\hat{\beta}) &= E(L^{-1}X^TY) = L^{-1}X^TE(Y) \\ &= L^{-1}X^TE(X\beta + \varepsilon) \\ &= L^{-1}X^T[X\beta + E(\varepsilon)] \\ &= L^{-1}X^TX\beta = \beta\end{aligned}$$

即
$$E(\hat{b}_i)=b_i \qquad (i=1\sim m)$$

2. $\hat{\beta}$的协方差矩阵为σ^2L^{-1},即

$$\begin{cases} D(\hat{b}_i) = \sigma^2 c_{ii}, \\ \mathrm{Cov}(\hat{b}_i,\hat{b}_j) = \sigma^2 c_{ij},(i,j=0,1,2,\cdots,m+1) \end{cases}$$

其中
$$L^{-1}=C=(c_{ij})_{(m+1)\times(m+1)}$$

证:记 $B=L^{-1}X^T$,则 $\hat{\beta}=BY$

$$\begin{aligned} \mathrm{Cov}(\hat{\beta},\hat{\beta}) &= E\{[BY-BE(Y)][BY-BE(Y)]^T\} \\ &= B\cdot E\{[Y-E(Y)][Y-E(Y)]^T\}B^T \\ &= B\cdot\sigma^2 I_n\cdot B^T \\ &= L^{-1}X^T\cdot\sigma^2 I_n\cdot(L^{-1}X^T)^T \\ &= \sigma^2 L^{-1} \end{aligned}$$

性质 2 告诉我们,用最小二乘法求出的诸回归系数 $b_0,b_1,b_2,\cdots,b_m$ 之间存在相关性,进一步可以证明

3. $\hat{\beta}$是 β 的最小方差线性无偏估计,即在所有线性无偏估计类中,有且只有 $\hat{\beta}$使其方差达到最小.

现在我们用式(5.9)和(5.10)计算例 1

$$L=X^TX=\begin{pmatrix} n & \sum_{i=1}^{n}x_{i1} & \sum_{i=1}^{n}x_{i2} \\ \sum_{i=1}^{n}x_{i1} & \sum_{i=1}^{n}x_{i1}^2 & \sum_{i=1}^{n}x_{i1}x_{i2} \\ \sum_{i=1}^{n}x_{i2} & \sum_{i=1}^{n}x_{i1}x_{i2} & \sum_{i=1}^{n}x_{i2}^2 \end{pmatrix}$$

$$=\begin{pmatrix} 10 & 441 & 147 \\ 441 & 19\,461 & 6\,485 \\ 147 & 6\,485 & 2\,173 \end{pmatrix}$$

$$X^TY=\begin{pmatrix} \sum_{i=1}^{n}y_i \\ \sum_{i=1}^{n}x_{i1}y_i \\ \sum_{i=1}^{n}x_{i2}y_i \end{pmatrix}=\begin{pmatrix} 272 \\ 12\,005 \\ 4\,013 \end{pmatrix}$$

L 的行列式

$$|L|=\begin{vmatrix} 10 & 441 & 147 \\ 441 & 19\,461 & 6\,485 \\ 147 & 6\,485 & 2\,173 \end{vmatrix}=1\,508$$

L 的伴随阵

$$L^{*}=\begin{pmatrix}233\ 528 & -4\ 998 & -882\\ -4\ 998 & 121 & -23\\ -882 & -23 & 129\end{pmatrix}$$

L 的逆阵

$$L^{-1}=\frac{L^{*}}{|L|}=\frac{1}{1508}\begin{pmatrix}233\ 528 & -4\ 998 & -882\\ -4\ 998 & 121 & -23\\ -882 & -23 & 129\end{pmatrix}$$

回归系数

$$\begin{pmatrix}b_0\\ b_1\\ b_2\end{pmatrix}=L^{-1}X^{-1}Y=\frac{1}{1\ 508}\begin{pmatrix}233\ 528 & -4\ 998 & -882\\ -4\ 998 & 121 & -23\\ -882 & -23 & 129\end{pmatrix}\begin{pmatrix}272\\ 12\ 005\\ 4\ 013\end{pmatrix}$$

$$=\begin{pmatrix}-13.819\ 63\\ 0.563\ 66\\ 1.099\ 47\end{pmatrix}$$

于是 y 对 x_1 及 x_2 的线性回归方程

$\hat{y}=-13.819\ 63+0.563\ 66x_1+1.099\ 47x_2$

用这个方程可预测管理费用，例如，劳动日数为 42 个单位，机器日平均开工台数为 13 个单位，则可预测管理费用为

$$\hat{y}=-13.819\ 63+0.563\ 66\times 42+1.099\ 47\times 13=24.147\ 2$$

四、多元线性回归方程的显著性检验

从上面的参数估计过程可以看出，对于一批观察数据

$$(y_i,x_{i1},x_{i2},\cdots,x_{im}),\qquad i=1\sim n$$

不论它们是否具有线性关系，总可以利用最小二乘法建立起多元线性回归方程

$$\hat{y}=\hat{b}_0+\hat{b}_1x_1+\hat{b}_2x_2+\cdots+\hat{b}_mx_m$$

但是 Y 与 $x_1,x_2,\cdots,x_m$ 是否确实存在相关关系呢？回归方程的效果如何呢？这就要进行“整个回归效果是否显著”的检验. 当 $b_1=b_2=\cdots=b_m=0$ 时，y 与 $x_1,x_2,\cdots,x_m$ 没有关系，回归模型没有意义，于是我们要检验 $H_0:b_1=b_2=\cdots=b_m=0$ 是否成立.

若 H_0 成立，则 $x_1,x_2,\cdots,x_m$ 对 y 没有影响；反之，若 H_0 不成立，则 $x_1,x_2,\cdots,x_m$ 对 y 有影响，此时 y 与 $x_1,x_2,\cdots,x_m$ 的线性关系显著，也称为整个回归效果显著. 但要注意，即使整个回归效果是显著的，y 也可能只与某几个 x_i 关系密切(相应的 b_i 显著不为零)，而与另几个 x_i 关系不密切(相应的 b_i 为零). 这就是说，多元线性回归除了首先要检验“整个回归是否显著”外，还要逐个检验每一个 b_i 是否为零，以便分辨出哪些 x_i 对 y 并无显著影响，最后，还要对各个 b_i 作出区间估计.

为了进行检验和区间估计，可以证明以下结论成立：

(1) $\frac{1}{\sigma^2}Q\sim\chi^2(n-m-1)$，且 Q 与 $\hat{b}_1,\hat{b}_2,\cdots,\hat{b}_m$ 独立

记
$$\bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i$$

$$l_{yy}=\sum_{i=1}^{n}(y_i-\bar{y})^2$$

则称 l_{yy} 为总变差或称为 y 的离差平方和. l_{yy} 可进行如下分解:

$$l_{yy}=\sum(y_i-\hat{y}_i)^2+\sum(\hat{y}_i-\bar{y})^2 ① =Q+U$$

这里 $Q=\sum(y_i-\hat{y}_i)$ 称为残差平方和.

$U=\sum(\hat{y}_i-\bar{y})^2$ 称为回归平方和.

记
$$s=\sqrt{\frac{Q}{n-m-1}}$$

称为剩余标准差或估计的标准差.

由于 l_{yy} 不变,当然希望 Q 越小越好,即 U 越大越好. 因此,定义复相关系数(complex correlation coefficient)

$$R=\sqrt{\frac{U}{l_{yy}}}=\sqrt{1-\frac{Q}{l_{yy}}}$$

当观察值 y_i 全都与回归值 $\hat{y}_i$ 吻合时, $Q=0, R=1$; 当 $\hat{y}_i=\bar{y}$ 时, $Q=l_{yy}, R=0$, 在一般情况下, R 的数值在 0 与 1 之间.

复相关系数 R 的定义,类似于两个变量时的相关系数的定义,但要注意,复相关系数 R 只取正值. 在两个变量时,有正相关与负相关之分,在多个变量时,就没有这一说了,所以复相关系数 R 只取正值.

(2) 在 $b_1=b_2=\cdots=b_m=0$ 的条件下

$$\frac{U}{\sigma^2}\sim\chi^2(m)$$

且 U 与 Q 独立,因此

$$F=\frac{U/m}{Q/(n-m-1)}=\frac{n-m-1}{m}\frac{R^2}{1-R^2}\sim F(m,n-m-1)$$

(3) $t_i=\dfrac{\hat{b}_i-b_i}{\sqrt{c_{ii}}\cdot s}\sim t(n-m-1), F_i=\dfrac{(\hat{b}_i-b_i)^2/c_{ii}}{Q/(n-m-1)}=\dfrac{(\hat{b}_i-b_i)^2}{c_{ii}^2 s^2}\sim F(1,n-m-1)$

$$i=1,2,\cdots,m$$

这里 c_{ii} 为 $C=L^{-1}$ 中第 i 个对角线元素.

利用上述几条结论,可进行下列检验、估计和预测.

(1) 回归显著性检验(F 检验)即整个回归效果是否显著性检验.

若整个回归效果不显著,即全部回归系数为零,因此,设原假设 $H_0: b_1=b_2=\cdots=b_m=0$. 若 H_0 为真,则

$$F=\frac{n-m-1}{m}\frac{U}{Q}=\frac{n-m-1}{m}\cdot\frac{R^2}{1-R^2}\sim F(m,n-m-1).$$

而且在 H_0 不成立时, F 值有变大的趋势,因此应取右侧否定域,故检验法是当 $F>F_\alpha(m,$

① 利用正规方程组,交叉乘积项求和为零

$n-m-1)$时拒绝原假设，认为回归效果显著；否则认为回归效果不显著.

(2) 单个回归系数为零的检验(t 检验)，即某个自变量是否对因变量有显著性影响的检验.

在多元回归分析中可能出现 y 与所有自变量的总体是有相关关系的，但 y 与某个特定的 x_i 则可能无关，即 x_i 对 y 并不起作用或者已被其它的 x_j 的作用所代替，为此设 m 个原假设

$$H_{0i}: b_i = 0, i = 1,2,\cdots,m.$$

若 H_{0i}为真，统计量

$$t_i = \frac{\hat{b}_i}{\sqrt{c_{ii}} \cdot s} \sim t(n-m-1), i = 1,2,\cdots,m$$

而当 H_{0i}不成立时，$|t_i|$有变大的趋势，因而应取双侧拒绝域，故当$|t_i|>t_{\frac{\alpha}{2}}(n-m-1)$时，否定 H_{0i}，即认为 x_i 对 y 是有作用的，若某几个 x_i 是有作用的，而另几个 x_j 是不起作用的，则应从回归方程中删除那些不起作用的自变量.

单个回归系数是否为零，也可以用 F 检验，即若 H_{0i}为真，统计量

$$F_i = \frac{\hat{b}_i^2}{c_{ii}^2 s^2} \sim F(1, n-m-1), i = 1,2,\cdots,m$$

故当 $F_i>F_\alpha(1,n-m-1)$时，拒绝原假设，即认为 x_i 对 y 的影响是显著的；否则认为 x_i 对 y 的影响是不显著的.

(3) 对 b_i 的区间估计.

由于 $$\frac{\hat{b}_i-b_i}{\sqrt{c_{ii}} \cdot s} \sim t(n-m-1)$$

因而 b_i 的 $1-\alpha$ 置信区间为

$$(\hat{b}_i - d_i, \hat{b}_i + d_i)$$

其中 $d_i=t_{\frac{\alpha}{2}}(n-m-1) \cdot \sqrt{c_{ii}} \cdot s$

(4) y_0 的 95%预测区间近似为$(\hat{y}_0-2S, \hat{y}_0+2S)$，其中

$$\hat{y}_0 = \hat{b}_0 + \hat{b}_1 x_{01} + \hat{b}_2 x_{02} + \cdots + \hat{b}_m x_{0m}$$

例 2　某种水泥在凝固时放出的热量 y(卡/克)与水泥中下列 4 种化学成份有关.

x_1: $3CaO \cdot Al_2O_3$ 的成份(%)

x_2: $3CaO \cdot SiO_2$ 的成份(%)

x_3: $4CaO \cdot Al_3O_3 \cdot Fe_3O_3$ 的成份(%)

x_4: $2CaO \cdot SiO_2$ 的成份(%)

现记录了 13 组数据，列在表 5-9 中，根据表中的数据：

(1)求 y 对 x_1, x_2, x_3, x_4 的线性回归方程；

(2)进行回归的显著性检验.

表 5-9

编号	x_1(%)	x_2(%)	x_3(%)	x_4(%)	y(卡/克)
1	7	26	6	60	78.5
2	1	29	15	52	74.3
3	11	56	8	20	104.3
4	11	31	8	47	87.6
5	7	52	6	33	95.9
6	11	55	9	22	109.2
7	3	71	17	6	102.7
8	1	31	22	44	72.5
9	2	54	18	22	93.1
10	21	47	4	26	115.9
11	1	40	23	34	83.8
12	11	66	9	12	113.3
13	10	68	8	12	109.4

下面介绍一下用 Math cai 软件求解例 2 的步骤：

1° 在 DOS 下打入可执行文件名__MULTREG〈回车〉；或者在 Math cai 子目录下打入 CAI，然后选择多元线性回归也可；

2° 按[F5]，建立一个新的数据文件.

出现询问：

列数(1—99)——输入变量总数 5〈回车〉

行数(1—999)——输入样本观察次数 13〈回车〉

进入数据编辑窗中，开始输入数据，输入数据的格式如下：

	x_1	……	x_m	Y
1.S1	x_{11}	……	x_{1m}	y_1
2.S2	x_{21}	……	x_{2m}	y_2
……	……	……	……	……
n.Sn	x_{n1}	……	x_{nm}	y_n

其中，$x_{ij}(i=1,2,\cdots,n,j=1,2,\cdots,m)$是第 i 次观测时，第 j 个自变量 x_j 的观测值，y_i $(i=1,2,\cdots,n)$是第 i 次观测时因变量 Y 的观测值，$S_i(i=1,2,\cdots,n)$是图像图示时，表示第 i 次观测的标记，如不写入，则在图像显示时，用数字 1，2，…表示各次观测.

数据输入完毕，可按[F2]存盘. 按[Esc]退出数据编辑窗，返回主菜单.（详见数据编辑窗中各键功能的说明）.

3° 在主菜单中，按[F9]，开始运行，进行多元线性回归运算.

自变量的个数________________4

样本观测次数________________13

变　　量	样本均值	样本方差	修正样本方差	样本标准差	修正样本标准差
x_1	7.461 54E+00	3.194 08E+01	3.460 26E+01	5.651 62E+00	5.882 39E+00
x_2	4.815 38E+01	2.235 15E+02	2.421 41E+02	1.495 04E+01	1.556 09E+01
x_3	1.176 92E+01	3.786 98E+01	4.102 56E+01	6.153 85E+00	6.405 13E+00
x_4	3.000 00E+01	2.586 15E+02	2.801 67E+02	1.608 15E+01	1.673 82E+01
Y	9.542 31E+01	2.089 05E+02	2.263 14E+02	1.445 35E+01	1.504 37E+01

显示样本相关阵?　　　　　　　　　　(Y/N)_____ Y

样本相关阵

	x_1	x_2	x_3	x_4	Y
x_1	1.000 000				
x_2	0.228 579	1.000 000			
x_3	−0.824 134	−0.139 242	1.000 000		
x_4	−0.245 445	−0.972 955	0.029 537	1.000 000	
y	−0.730 717	0.816 253	−0.534 671	−0.821 305	1.000 000

(1) $Y=B_0+B_1x_1+B_2x_2+\cdots+B_mx_m$

(2) $Y=B_1x_1+B_2x_2+\cdots+B_mx_m$

(3) $Y=B_0*x_1{}^{B_1}*x_2{}^{B_2}*\cdots*x_m{}^{B_m}$

(4) $Y=x_1{}^{B_1}*x_2{}^{B_2}*\cdots*x_m{}^{B_m}$

(0) 退出

键入 1,2,3,4 或 0 _______ 1

$Y=B_0+B_1x_1+B_2x_2+\cdots+B_mx_m$　　$(m=4)$

$B_0=6.240\,536\,9E+01$

$B_1=1.551\,102\,6E+00$

$B_2=5.101\,675\,8E-01$

$B_3=1.019\,094\,0E-01$

$B_4=-1.440\,610\,3E-01$

显示图像?　　　　　　　　　　(Y/N)_____ N

Y 的样本观测值,估计值和残差

	Y	$\hat{Y}$	$Y-\hat{Y}$		Y	$\hat{Y}$	$Y-\hat{Y}$
1	7.850 0E+01	7.849 5E+01	4.76E−03	8	7.250 0E+01	7.567 5E+01	−3.17E+00
2	7.430 0E+01	7.278 9E+01	1.51E+00	9	9.310 0E+01	9.172 2E+01	1.38E+00
3	1.043 0E+02	1.059 7E+02	−1.67E+00	10	1.159 0E+02	1.156 2E+02	2.82E−01
4	8.760 0E+01	8.932 7E+01	−1.73E+00	11	8.380 0E+01	8.180 9E+01	1.99E+00
5	9.590 0E+01	9.564 9E+01	2.51E−01	12	1.133 0E+02	1.123 3E+02	9.73E−01

6　1.092 0E+02　1.052 7E+02　3.93E+00　　13　1.094 0E+02　1.116 9E+02　−2.29E+00

7　1.027 0E+02　1.041 5E+02−1.45E+00

残差平方和(剩余平方和)=4.786 363 9E+01

估计的标准差=2.446 008 0E+00

多重相关系数(复相关系数)=0.991 148 6

临界值

F(4,8)分布统计量　F=111.479　$F_{0.05}$=3.838　$F_{0.01}$=7.006

将计算结果存入数据文件?　　(Y/N)_____

从上述计算结果易见,回归方程

$$\hat{y}=62.4054+1.5511x_1+0.5102x_2+0.1019x_3-0.1441x_4$$

且统计量 $F=111.479>F_{0.05}(4,8)=3.838$,$F>F_{0.01}(4,8)=7.006$,所以,回归效果是极高度显著的.

第三节　多项式回归模型及广义线性回归模型

一、多项式回归模型

由于任一连续函数都可以用多项式函数逼近,所以在实际问题中常用多项式函数进行回归.

例1　在对树龄 x 和树的高度 y 的观测中得如下资料见表 5-9.

表 5-9

x_i	2	3	4	5	6	7	8	9	10	11
y_i	5.6	8	10.4	12.8	15.3	17.8	19.9	21.4	22.4	23.2

我们选取一元二次多项式

$$Y=b_0+b_1x+b_2x^2+\varepsilon,\varepsilon\sim N(0,\sigma^2) \tag{5.12}$$

来拟合它,试求其回归方程

解　令 $x_1=x,x_2=x^2$,则模型(5.12)可改写成

$$Y=b_0+b_1x_1+b_2x_2+\varepsilon,\varepsilon\sim N(0,\sigma^2)$$

这是一个二元线性回归模型.

仿多元线性回归模型的计算方法,可解得

$\hat{b}_0=-1.33,\hat{b}_1=3.46,\hat{b}_2=-0.11$

于是 Y 关于 x 的回归方程为

$$\hat{y}=-1.33+3.46x-0.11x^2$$

为了进行显著性检验,设原假设 $H_0:b_1=b_2=0$,而

$$F=\frac{n-m-1}{m}\frac{U}{Q}$$

$$= \frac{10-2-1}{2} \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

$$= 997.9$$

而查表知 $F_{0.05}(2,7)=4.74$，即 $F>F_{0.05}(2,7)$，故拒绝 H_0，即认为回归效果是显著的.

类似于模型(5,5)，由自变量多项式和随机项组成的回归模型称为多项式回归模型，它的一般形式为

$$Y = b_0 + b_1x + b_2x^2 + \cdots + b_mx^m + \varepsilon, \varepsilon \sim N(0,\sigma^2) \tag{5.13}$$

初看模型(5.13)不是线性回归，因自变量中含有幂函数，但由于未知参数 $b_0,b_1,\cdots,b_m$ 都是线性出现的，因此，令

$$x_1 = x, x_2 = x^2, \cdots, x_m = x^m$$

则模型(5.13)就变成为多元线性回归模型：

$$Y = b_0 + b_1x_1 + b_2x_2 + \cdots + b_mx_m + \varepsilon \tag{5.14}$$

从而多项式回归模型可以用多元线性回归模型的计算公式和检验方法.

多项式回归还有许多推广的形式，例如：

(1) $y=b_0+b_1x+b_2x^2+\cdots+b_mx^m+\frac{c}{x}$

(2) $y=b_0+b_1x+b_2x^2+\cdots+b_mx^m+C\mathrm{Ln}^x$

(3) $y=Exp(b_0+b_1x+b_2x^2+\cdots+b_mx^m)$

(4) $y=Exp(b_0+b_1x+b_2x^2+\cdots+b_mx^m+\frac{c}{x})$

(5) $y=Exp(b_0+b_1x+b_2x^2+\cdots+b_mx^m)x^c$

上述模型的共同特点是未知参数都是以线性形式出现，所以都可以采用恒等变换像模型(5.13)化为模型(5.14)一样化为多元线性回归模型.

二、广义线性回归模型

广义线性回归模型的一般形式为

$$y = f(b_0 + b_1F_1(x_1,x_2,\cdots,x_m) + \cdots + b_pF_p(x_1,x_2,\cdots,x_m)) \tag{5.15}$$

其中：$y=f(y_0)$ 是一个不含未知数参数的一元函数，且有反函数：

$$y_0 = g(y)$$

$F_j(x_1,x_2,\cdots,x_m)(j=1,2,\cdots,p)$ 是 $x_1,x_2,\cdots,x_m$ 的不含未知参数的多元函数.

广义线性回归模型的回归系数的确定主要是从自变量 $x_1,x_2,\cdots,x_m$ 和因变量 y 的 n 组观察值

$$(x_{i1},x_{i2},\cdots,x_{im},y_i), i = 1,2,\cdots,n$$

出发，用最小二乘法求出 $b_0,b_1,\cdots,b_p$ 的估计 $\hat{b}_0,\hat{b}_1,\cdots,\hat{b}_p$，使得

$$Q = \sum_{i=1}^{N}[g(y_i) - (b_0 + b_1F_1(x_{i1},x_{i2},\cdots,x_{im}) + \cdots + b_pF_p(x_{i1},x_{i2},\cdots,x_{im}))]^2$$

达到最小.

此时也就是令

$$\begin{cases} Y = g_0(y), \\ t_1 = F_1(x_1, x_2, \cdots, x_m) \\ \cdots \quad \cdots \quad \cdots \\ t_p = F_p(x_1, x_2, \cdots, x_m) \end{cases}$$

则

$$Y = b_0 + b_1 t_1 + \cdots + b_p t_p \tag{5.16}$$

这样就把广义线性回归模型化为多元线性回归模型.

例如 $y = a_0 + a_1 \sin\omega_1 x + b_1 \cos\omega_1 x + \cdots + a_m \sin\omega_m x + b_m \cos\omega_m x$

这里 $\omega_1, \omega_2, \cdots, \omega_m$ 为已知的常数,就属于广义线性回归模型.

三、最优回归模型的选择

很多实际问题中,例如气象预报、地震预报等,自变量个数多得惊人,过多的自变量不仅给数据的收集、计算带来很大麻烦,而且多个自变量之间往往会产生线性相关或近似线性相关关系.这种相关关系,称为"复共线性".复共线性的存在,会使得回归分析的误差增大,得到不合理的结果,甚至使计算溢出,无法求得回归分析问题的解.所以,我们总是希望从众多自变量中挑选较少的一些变量,建立"最优"的回归模型.所谓"最优"的回归模型有两个方面的含义:一方面指方程中要包含所有的有显著作用的自变量;另一方面,希望自变量的个数尽可能地少,因为自变量过多,使用起来不方便,而且引入了无意义的量,反而会使回归方程的稳定性降低.

为了解决这些问题,就产生了一种重要的新的方法,即逐步回归法.它是一种能避免由复共线性产生的不良影响的回归分析方法.其基本思想是:不取全部自变量作回归,而是从一个含自变量的方程出发,通过逐步引入和删除,从全部自变量中选取一部分自变量作回归,既要使残差平方和尽量小,又要避免引入过多的自变量,以致产生复共线性现象.

下面简单介绍一下数学原理.

设有 $m+1$ 个变量 $y_1, x_1, x_2, \cdots, x_m$ 的 n 次观察值,用回归模型($p \leqslant m$)

$$y = b_0 + b_1 x_1 + \cdots + b_p x_p$$

来表示它,得回归方程

$$\hat{y} = \hat{b}_0 + \hat{b}_1 x_1 + \cdots + \hat{b}_p x_p$$

及回归平方和 U,再用回归模型

$$y = b_0 + b_1 x_1 + \cdots + b_{i-1} x_{i-1} + b_{i+1} x_{i+1} + \cdots + b_p x_p$$

来表示它,得回归方程

$$\hat{y} = \hat{b}_0^* + \hat{b}_1^* x_1 + \cdots + \hat{b}_{i-1}^* x_{i-1} + \hat{b}_{i+1}^* x_{i+1} + \cdots + \hat{b}_p^* x_p$$

和回归平方和 U_i,则

$$S_i = U - U_i$$

称为偏回归平方和.若 x_i 可用其余自变量线性表示,则 U 与 U_i 应当差不多,因而 S_i 应当很小.所以当从 F 检验中发现 S_i 很小时,我们就可以从 P 个自变量中删除它.类似地,我们也可以检验 $x_{p+1}, \cdots, x_m$ 中的自变量是否应当添到回归模型中去.

逐步回归的具体做法是：事先给定一个引入变量的 F 水平界限 F_{in} 和一个删除变量的 F 水平界限 F_{out}，首先在所有尚未引进方程的自变量中，找出 F_i 值最大的一个，如果 $F_i > F_{in}$ 就引入自变量 x_i，然后在所有已引进方程的自变量中，找出 F_i 值最小的一个，如果 $F_i \leqslant F_{out}$，就删除自变量 x_i. 这样逐步引入、删除，直到方程外没有一个自变量的 F_i 大于 F_{in}，方程内没有一个自变量的 F_i 小于 F_{out} 为止.

在逐步回归中，检验作用显著的临界值 F 像一个门坎，它的值取大些，进入回归方程的变量就会少些，反之就会多些.

在实际计算中，为了避免出现一个自变量反复引入、删除、循环不止的现象，事先给定的水平界限必须满足 $F_{in} \geqslant F_{out} \geqslant 0$. 另外，为了避免引入自变量后产生复共线性，使回归分析计算误差增大或溢出，还必须事先给定一个容许值的水平界限 $Tol(0 < Tol \leqslant 0.1)$，在逐步回归过程中，如果引进一个自变量时，计算公式中分母的绝对值小于 TOl 就不引进这个自变量.

例 逐步回归在生猪生产决策中的应用.

影响生猪生产的原因很多，为寻找影响生猪生产的原因，现搜集整理了某市 1980～1990 年共 11 年的粮食产量、生猪外调量、鲜鱼产量、家禽产量、猪肉销售价格和生猪出栏量的统计数据(见表 5-10)，试采用逐步回归的方法，研究影响生猪出栏量的多元线性回归模型，并据此对未来的生猪生产进行预测.

表 5-10

年份 \ 指标	粮食产量(万吨) x_1	猪外调量(千头) x_2	鲜鱼产量(百吨) x_3	家禽产量(万只) x_4	猪肉售价(元/公斤) x_5	生猪出栏(万头) Y
1980	215.52	190.94	132.77	770	1.60	210.29
1981	219.91	270.42	153.30	834	1.607	222.09
1982	229.53	230.72	120.73	817	1.627	218.81
1983	258.44	250.58	188.05	906	2.014	237.02
1984	243.25	305.76	241.72	1 161	1.978	263.19
1985	211.57	358.45	255.38	1 159	2.262	277.74
1986	243.43	426.07	255.85	1 227	2.437	272.42
1987	239.79	414.77	304.64	1 184	2.827	313.25
1988	245.94	345.53	305.17	1 148	4.185	321.91
1989	253.17	380.18	353.96	1 358	4.547	330.09
1990	232.08	333.60	335.88	1 289	4.344	335.92

1. 建立数学模型

逐步回归就是从与 Y(因变量)有关的变量中选取对 Y 有显著影响的变量来建立回归模型. 为此，选取生猪出栏量作因变量 Y，用直接影响生猪生产的粮食产量 x_1，猪外调量

x_2，鲜鱼产量 x_3，家禽产量 x_4，猪肉售价 x_5 作为自变量.

假设上一年度各自变量的变化影响下一年度的

$$Y_t = f(x_{i(t-1)}) \tag{5.17}$$

t——年份，$i=1,2,3,4,5$

对于某些变量，这一点是显然的.例如 x_1（粮食产量），上一年的丰歉程度，对下一年的粮食消耗至关重要，若其对生猪饲养有影响，自然应是上一年而不是本年度的粮食产量在起主要作用，其它因素也是如此.

为寻找出影响生猪出栏量的关键变量，设计多元线性逐步回归模型来拟合生猪出栏量的关键变量，设计多元线性逐步回归模型来拟合生猪出栏量与各自变量 x_i 的数量关系，其模型形式如下：

$$Y_t = b_0 + \sum_{i=1}^{5} b_i x_{i(t-1)} \tag{5.18}$$

式中 Y_t——第 t 年的生猪出栏量；

$x_{1(t-1)}$——第 $(t-1)$ 年的粮食产量；

$x_{2(t-1)}$——第 $(t-1)$ 年的生猪外调量；

$x_{3(t-1)}$——第 $(t-1)$ 年的鲜鱼产量；

$x_{4(t-1)}$——第 $(t-1)$ 年的家禽产量；

$x_{5(t-1)}$——第 $(t-1)$ 年的猪肉售价；

b_i——待估计参数，$i=1,2,3,4,5$.

2. 应用逐步回归求解模型

2.1 先考虑一个自变量

人们普遍认为粮食产量 x_1 是生猪生产的气候表，故首先引入 x_1，此时(5.18)变为

$$y = b_0 + b_1 x_1$$

按最小二乘法，求得回归方程

$$\hat{y} = -135.82 + 1.76x_1$$

$R^2=0.40$，这表明，对于生猪出栏量来说，粮食产量只解释了总变差 40%.

以上结果并不象人们一般地认识那样，生猪生产最主要地取决于粮食生产状况，严格地说，可以认为当粮食减少到某一临界范围时，即只能满足人们生活的基本必需时，才有可能对生猪生产产生决定性的影响.

2.2 为简便起见，同时引入其余 5 个变量，(5.18)变为

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_5$$

下面利用 Math cai 软件求解该问题.

1° 在 Math cai 子目录下打入 cai，进入逐步回归，选择操作练习（回车）；

2° 出现询问：

回归方程 $Y=f(B_0+B_1Z[1]+\cdots+B_pZ[p])$

回归方程中变量 Z 的个数 $p(1-24)$______

本题输入回归方程中自变量 x 的个数 5（回车），进入表达式输入窗口：

回归方程　$Y=f(B_0+B_1Z[1]+\cdots+B_pZ[p])$

写入函数表达式　$Z[j]:=Fj(X[1],\cdots,X[m])$;　$j=1,\cdots,p$

```
 
 
 
 
                    1:1
```

表达式中,可使用实型中间变量:r,s,t,u,v,w　可使用整型中间变量:i,j,k,l

可使用标准函数:$\mathrm{Sqr}(x)=x^2$　$\mathrm{Sqrt}(x)=\sqrt{x}$　$\mathrm{Abs}(x)=|x|$　$\mathrm{Exp}(x)=e^x$　$\mathrm{Ln}(x)$　$\mathrm{Sin}(x)$　$\mathrm{Cos}(x)$　$\mathrm{Arctan}(x)=\mathrm{Arctg}x$　$Pi=\pi$　$\max(x_1,x_2)$　$\min(x_1,x_2)$

在上面表达式窗口中输入函数表达式后变为:

回归方程　$Y=f(B_0+B_1Z[1]+\cdots+B_pZ[p])$

写入函数表达式　$z[j]:=F_j(X[1],\cdots,X[m])$;　$j=1,\cdots,p$

```
Z[1]:=X[1];
Z[2]:=X[2];
Z[3]:=X[3];
Z[4]:=X[4];
Z[5]:=X[5];
                    6:1
```

按[Esc]退出,进入下一个窗口:

回归方程　$Y=f(B_0+B_1Z[1]+\cdots+B_pZ[p])$

写入函数表达式　$Y:=f(Y_0)$;

```
 
 
 
 
                    1:1
```

在上面窗口中输入函数表达式:

$$Y:=Y_0;$$

按[Esc]退出,进入下一个窗口.

回归方程　$Y=f(B_0+B_1Z[1]+\cdots+B_pZ[p])$

写入函数表达式　　$Y_0:=g(Y)$;{$Y=f(Y_0)$的反函数}

1,1

在上面窗口中输入 $Y=Y_0$ 的反函数表达式：

$$Y_0:=Y;$$

按[Esc]退出.

3° 按[F5]建立新的数据文件,(若已有数据文件,按[F3]打开既可).

出现询问:列数(1—99)——输入变量总数 6(回车)

行数(1—999)——输入样本观测次数 11(回车)

进入数据编辑窗口中,输入数据格式如下:

	x_1	x_2	x_3	x_4	x_5	Y
1.	215.52	190.94	132.77	770	1.60	210.29
2.	219.91	207.42	153.30	834	1.607	222.09
3.	229.53	230.72	120.73	817	1.627	218.81
4.	258.44	250.58	188.05	906	2.104	237.02
5.	243.25	305.76	241.72	1 161	1.978	263.19
6.	211.57	358.45	255.38	1 159	2.262	277.74
7.	243.43	426.07	255.85	1 227	2.437	272.42
8.	239.79	414.77	304.64	1 184	2.827	313.25
9.	245.94	345.53	305.17	1 148	4.185	321.91
10.	253.17	380.18	353.96	1 358	4.547	330.09
11.	232.08	333.60	335.88	1 289	4.344	335.92

按[F2]将数据文件存盘;数据文件名______________ B:STEPREG.DAT 然后按[Esc]退出.

4° 按[F9]运行,显示结果如下:

数据文件名______________ B:STEPREG.DAT

变量 X 的个数______________ 5

变量 Z 的个数______________ 5

样本观测次数______________ 11

变量	样本均值	样本方差	修正样本方差	样本标准差	修正样本标准差
Z1	2.356 94E+02	2.136 94E+02	2.350 64E+02	1.461 83E+01	1.533 18E+01
Z2	2.188 20E+02	5.296 47E+03	5.826 11E+03	7.277 68E+01	7.632 90E+01
Z3	2.406 77E+02	6.093 66E+03	6.703 03E+03	7.806 19E+01	8.187 20E+01

Z4	1.077 55E+03	3.880 10E+04	4.268 11E+04	1.969 80E+02	2.065 94E+02
Z5	2.675 27E+00	1.194 33E+00	1.313 76E+00	1.092 85E+00	1.146 20E+00
Y*	2.731 57E+02	1.987 08E+03	2.185 79E+03	4.457 67E+0.	4.675 24E+01

显示样本相关阵?　　　　　　　　　　　　(Y/N)______ Y

样本相关阵

	x_1	x_2	x_3	x_4	Y_0	
x_1	1.000 000					
x_2	0.324 306	1.000 000				
x_3	0.432 378	0.809 721	1.000 000			
x_4	0.406 297	0.864 818	0.955 130	1.000 000		
x_5	0.430 211	0.558 086	0.892 427	0.781 377	1.000 000	
Y_0	0.388 855	0.765 355	0.983 090	0.909 935	0.927 615	1.000 000

按任意键继续……

(1) $Y=f(B_0+B_1Z[1]+B_2Z[2]+\cdots+B_pZ[p])$

(2) $Y=f(B_1Z[1]+B_2Z[2]+\cdots\cdots+B_pZ[p])$

(0)退出

键入 1,2 或 0 ______

输入 1(也就是选择回归方程为 $Y=B_0+B_1x_1+B_2x_2+B_3x_3+B_4x_4+B_5x_5$)

显示

引入变量的 F 水平界限 Fin __________ 1

删除变量的 F 水平界限 Fout __________ 0.5

容许值的水平界限 Tol(0—0.1)__________ 0.0001

(上述引入变量的 F 水平界限 Fin 中的 1,删除变量的 F 不平界限 Fout 中的 0.5,容许值的水平界限 Tol 中的 0.0001 由操作者任意输入)

$Y=f(B_0+B_1Z[1]+B_2Z[2]+\cdots+B_pZ[P])$　$(P=5)$

$B_0=2.7297545E+02$		
$B_1=0$	Tol=1.000 000	$F_1=1.636\ 761\ 6E+00$
$B_2=0$	Tol=1.000 000	$F_2=1.295\ 761\ 5E+01$
$B_3=0$	Tol=1.000 000	$F_3=2.609\ 084\ 3E+02$
$B_4=0$	Tol=1.000 000	$F_4=4.401\ 160\ 4E+01$
$B_5=0$	Tol=1.000 000	$F_5=5.281\ 352\ 6E+01$

$Y=f(B_0+B_1Z[1]+B_2Z[2]+\cdots+B_pZ[P])$　$(P=5)$

$B_0=1.3708771E+02$		
$B_1=0$	Tol=0.813 050	$F_1=3.313\ 217\ 3E+01$

$B_2=0$	Tol=0.344 352	F_2=5.824 148 3E−01
B_3=5.646 056 4E−01		F_3=2.609 084 3E+02
$B_4=0$	Tol=0.087 726	F_4=2.901 815 8E+00
$B_5=0$	Tol=0.205 807	F_5=4.080 838 0E+00

$Y=f(B_0+B_1Z[1]+B_2Z[2]+\cdots+B_pZ[P])$ ($P=5$)

B_0=1.400 417 0E+02		
$B_1=0$	Tol=0.796 463	F_1=9.814 565 9E−01
$B_2=0$	Tol=0.206 920	F_2=2.060 545 1E−01
B_3=4.448 816 5E−01		F_3=4.475 091 4E−01
$B_4=0$	Tol=0.061 821	F_4=6.569 261 5E−01
B_5=9.637 149 4E+00		F_5=4.080 838 0E+00

显示图像？　　　　(Y/N)______ Y

Y 的样本观测值、估计值和残差

	Y	$\hat{Y}$	$Y-\hat{Y}$		Y	$\hat{Y}$	$Y-\hat{Y}$
1	2.102 9E+02	2.145 3E+02	−4.24E+00	7	2.724 2E+02	2.773 5E+02	−4.93E+00
2	2.220 9E+02	2.237 3E+02	−1.64E+00	8	3.132 5E+02	3.028 1E+02	1.04E+01
3	2.188 1E+02	2.094 3E+02	9.38E+00	9	3.219 1E+02	2.161 4E+02	5.77E+00
4	2.370 2E+02	2.439 8E+02	−6.96E+00	10	2.200 9E+02	3.413 3E+02	−1.12E+01
5	2.631 9E+02	2.666 4E+02	−3.45E+00	11	3.359 2E+02	3.313 3E+02	4.59E+00
6	2.777 4E+02	2.754 5E+02	2.29E+00				

残差平方和(剩余平方和)=4.881 000 7E+02
估计的标准差=9.880 284 1E+00
多重相关系数(复相关系数)=0.988 897 9
将计算结果存入数据文件？　　　　(Y/N)______
若输入 Y，可将已计算的结果存入一个新的数据文件；
若输入 N，又返回到下列状态：
(1)$Y=f(B_0+B_1Z[1]+B_2Z[2]+\cdots+B_pZ[p])$
(2)$Y=f(B_1Z[1]+B_2Z[2]+\cdots+B_pZ[p])$
(0)退出
键入 1，2，或 0 ______

键入 1 或 2 后，可重新输入引入变量的 F 水平界限 Fin，删除变量的 F 水平界限 Fout 及容许值的水平界限 Tol，比如，分别输入 0.5，0.1，0.000 01，显示结果如下：

引入变量的 F 水平界限　Fin ______________ 0.5
删除变量的 F 水平界限　Fout ______________ 0.1
容许值的水平界限　Tol(0−0.1)______________ 0.000 01

$$Y=f(B_0+B_1Z[1]+B_2Z[2]+\cdots+B_pZ[p])\quad(p=5)$$

$B_0=2.7297545E+02$		
$B_1=0$	Tol=1.000 000	$F_1=1.636\ 761\ 6E+00$
$B_2=0$	Tol=1.000 000	$F_2=1.295\ 761\ 5E+02$
$B_3=0$	Tol=1.000 000	$F_3=2.609\ 084\ 3EE+01$
$B_4=0$	Tol=1.000 000	$F_4=4.401\ 160\ 4E+01$
$B_5=0$	Tol=1.000 000	$F_5=5.281\ 352\ 6E+01$

$$Y=f(B_0+B_1Z[1]+B_2Z[2]+\cdots+B_pZ[P])\quad(P=5)$$

$B_0=1.370\ 877\ 1E+02$		
$B_1=0$	Tol=0.813 050	$F_1=3.313\ 217\ 3E+01$
$B_2=0$	Tol=0.344 352	$F_2=5.824\ 148\ 3E-01$
$B_3=5.646\ 056\ 4E-01$		$F_3=2.609\ 084\ 3E+02$
$B_4=0$	Tol=0.087 726	$F_4=2.901\ 815\ 8E+00$
$B_5=0$	Tol=0.205 807	$F_5=4.080\ 838\ 0E+00$

$$Y=f(B_0+B_1Z[1]+B_2Z[2]+\cdots+B_pZ[P])\quad(P=5)$$

$B_0=1.400\ 417\ 0E+02$		
$B_1=0$	Tol=0.796 463	$F_1=9.814\ 565\ 9E-01$
$B_2=0$	Tol=0.206 920	$F_2=2.060\ 545\ 1E-01$
$B_3=4.448\ 816\ 5E-01$		$F_3=4.475\ 091\ 4E-01$
$B_4=0$	Tol=0.061 821	$F_4=6.569\ 261\ 5E-01$
$B_5=9.637\ 149\ 4E+00$		$F_5=4.080\ 838\ 0E+00$

$$Y=f(B_0+B_1Z[1]+B_2Z[2]+\cdots+B_pZ[P])\quad(P=5)$$

$B_0=1.789\ 626\ 4E+02$		
$B_1=-1.790\ 496\ 4E-01$		$F_1=9.814\ 565\ 9E-01$
$B_2=0$	Tol=0.206 315	$F_2=2.528\ 762\ 7E-01$
$B_3=4.508\ 963\ 1E-01$		$F_3=4.548\ 438\ 0E+01$
$B_4=0$	Tol=0.061 573	$F_4=5.494\ 539\ 7E-01$
$B_5=1.031\ 997\ 5E+01$		$F_5=4.573\ 515\ 8E+00$

$$Y=f(B_0+B_1Z[1]+B_2Z[2]+\cdots+B_pZ[P])\quad(P=5)$$

$B_0=1.947\ 151\ 5E+02$		
$B_1=-1.702\ 633\ 6E-01$		$F_1=8.270\ 454\ 1E-01$
$B_2=0$	Tol=0.190 329	$F_2=5.182\ 527\ 7E-01$
$B_3=5.691\ 900\ 1E-01$		$F_3=1.071\ 171\ 1E+01$
$B_4=-3.696\ 683\ 8E-02$		$F_4=5.494\ 539\ 7E-01$
$B_5=7.912\ 429\ 6E+00$		$F_5=1.766\ 663\ 2E+00$

$$Y=f(B_0+B_1Z[1]+B_2Z[2]+\cdots+B_pZ[P])\quad(P=5)$$

$B_0=1.941\ 192\ 0E+02$		
$B_1=-1.755\ 902\ 3E-01$		$F_1=8.078\ 198\ 3E-01$

$B_2=5.763\ 204\ 2E-02$ $\qquad$ $F_2=5.182\ 527\ 7E-01$

$B_3=5.215\ 398\ 1E-01$ $\qquad$ $F_3=7.298\ 839\ 2E+00$

$B_4=-4.781\ 662\ 6E-02$ $\qquad$ $F_4=7.799\ 868\ 3E-01$

$B_5=1.038\ 561\ 0E+01$ $\qquad$ $F_5=2.142\ 909\ 1E+00$

显示图像? (Y/N)_____

Y 的样本观测值，估计值和残差

	Y	$\hat{Y}$	$Y-\hat{Y}$		Y	$\hat{Y}$	$Y-\hat{Y}$
1	2.102 9E+02	2.163 2E+02	−6.03E+00	7	2.724 2E+02	2.760 1E+02	−3.59E+00
2	2.220 9E+02	2.278 5E+02	−5.76E+00	8	3.132 5E+02	3.075 5E+02	5.70E+00
3	2.188 1E+02	2.079 1E+02	1.09E+01	9	3.219 1E+02	3.185 8E+02	3.33E+00
4	2.370 2E+02	2.397 9E+02	−2.77E+00	10	3.300 9E+02	3.384 7E+02	−8.38E+00
5	2.631 9E+02	2.601 2E+02	3.07E+00	11	3.359 2E+02	3.312 5E+02	4.67E+00
6	2.777 4E+02	2.788 9E+02	1.15E+00				

残差平方和(剩余平方和)$=3.553\ 360\ 9E+02$

估计的标准差$=8.430\ 137\ 5E+00$

多重相关系数(复相关系数)$=0.991\ 930\ 0$

将计算结果存入数据文件? (Y/N)_____

从上述计算结果可见:

若取引入变量的 F 水平界限 Fin=1，删除变量的 F 水平界限 Fout=0.5，则逐步回归方程为

$$\hat{y}=140.041\ 7+0.444\ 88x_3+9.637\ 149x_5$$

从逐步回归方程可见，影响生猪出栏量 Y 的主要因素是猪肉售价，其次是猪肉的替代品鲜鱼产量.

将 1990 年的各变量数据代入上述逐步回归方程，可预测 1991 年的生猪产量

$$Y_{1991}=331.33(\text{万头})$$

若取引入变量的 F 水平界限 Fin=0.5，删除变量的 F 水平界限 Fout=0.1，则逐步回归方程为:

$$\hat{y}=194.119\ 2-0.175\ 59x_1+0.057\ 632x_2+0.521\ 539\ 8x_3$$
$$-0.047\ 816\ 6x_4+10.385\ 61x_5$$

输入不同的 Fin 和 Fout，可以得到不同的逐步回归方程.

第四节 方差分析

方差分析是分析试验数据的一种重要方法. 在科学试验和生产实践中，影响一事物的因素往往是很多的，有的因素影响可能大些，有的因素影响可能小些，那么在诸多因素中怎么来确定那些因素对产品指标有显著影响?例如，在工业生产中，产品质量指标受原料、

设备、技术水平等因素的影响;在农业生产中,农作物的产量受种子、水分、肥料等因素的影响. 为了增加产量、提高质量,就要从诸多因素中找出哪个因素或哪些因素有显著的影响,以及在怎样的状态下其影响最大,这就需要进行试验. 方差分析就是根据试验的数据进行分析,鉴别各有关因素对试验结果影响的一种有效方法. 为什么叫方差分析呢? 因为它把一组数据的总变差分解为若干种成分,与这些成分中每一种成分相联系的是某一特殊变差的来源,而在分析中有可能确定每一特殊变差对总变差贡献的大小,所以这种分析方法称为方差分析. 它不仅在各项具体工作中,如气象预报、农业试验、工业、生物学、医学等领域有广泛的应用,而且它的思想已经渗透到了数理统计的许多方法之中. 学习方差分析不仅要掌握具体方法,还要领会它的统计思想.

方差分析的内容很多,本节将主要介绍单因素试验方差分析和双因素试验方差分析.

一、单因素试验方差分析

下面通过实例来说明方差分析的基本思想.

例 1 某电光源开发公司用四种不同配料方案制成的灯丝生产了四批灯炮. 在每一批灯炮中随机地抽若干只作寿命试验,得数据如表 5-11 所示. 问灯丝的不同配料方案对灯炮寿命有无显著影响?

如果这四种配料方案对灯炮的使用寿命没有显著影响,我们就可以从中选取一种最经济的配料方案;如果有显著影响,则希望知道哪一种配料方案比较好.

表 5-11

试验次数 / 使用寿命 / 水平	1	2	3	4	5	6	7	8
A_1	160 0	161 0	165 0	168 0	170 0	170 0	178 0	
A_2	150 0	164 0	140 0	170 0	175 0			
A_3	164 0	155 0	160 0	162 0	164 0	160 0	1740	180 0
A_4	151 0	152 0	153 0	157 0	164 0	168 0		

例 2 考察实验室一小时内,在不同电流强度下得到的电解铜的纯度,对每种电流强度各做了 5 次试验,分别测得其含杂质率数据见表 5-12. 试判断电流强度对电解铜的杂质率是否有显著影响?

表 5-12

杂质率(%) / 样品号 / 水平		1	2	3	4	5
A_1	10	1.7	2.1	2.2	2.1	1.9
A_2	15	2.1	2.2	2.0	2.2	2.1
A_3	20	1.5	1.3	1.8	1.4	1.7
A_4	25	1.9	1.9	2.2	2.3	2.0

这两个例子和以前回归分析的问题不同. 首先它只是考察某种因素(配料方案、电流)在一系列试验中对产品的某个指标(寿命、杂质率)的影响是否显著,而不要求建立回归方程;其次,这些因素只是一种属性,不是定量的(如配料方案),即或是定量的(如电流)但其数值也不作为回归中变量的观察值,而只是代表一种试验方案;最后,当因素确定后,可以作重复的试验.

在试验中,我们将所要观察的数量指标称为试验指标,如例 1 中的寿命,例 2 中的杂质率;影响试验指标的条件称为因素(或因子),用大写字母 A、B、C 表示,如例 1 中的配料方案,例 2 中的电流;因素所处的状态称为水平,如例 1、例 2 中的 A_1、A_2、A_3、A_4,水平常以 A_2、A_2、…、B_1、B_2…表示. 由于因素水平的不同而引起的试验指标的差异,称为条件误差. 而在同一水平下,由许多我们不便于控制或者根本无法控制的偶然因素所引起的试验指标的差异,则称为试验误差或称为随机误差.

如果一项试验中,只有一个因素变化,其它因素保持不变的话,称为单因素试验. 单因素试验的方差分析就是分析单因素试验中我们所关注的因素的变化对试验结果是否有显著的影响.

对于例 1 来说,进行试验时,除制成灯丝的配料方案这一因素外,其它的因素如生产条件都应保持不变. 对于每一个确定的水平,虽然配料不变,生产条件相同,但灯泡的使用寿命是一个随机变量. 所以,四种配料方案下的使用寿命仍看成是来自四个总体 X_1,X_2,X_3,X_4. 要确定这四种不同的配料方案对灯泡的使用寿命有无影响,也就是要确定这四个总体是否具有相同的分布. 假定用这四种不同的配料方案制成的灯丝生产的灯泡的寿命都近似地服从正态分布,且具有相同的方差,即

$$X_j \sim N(\mu_j,\sigma^2) \qquad j=1,2,3,4$$

则分析用四种不同的配料方案制成的灯丝是否会显著地影响灯泡的使用寿命,实际上就是检验具有相同方差的四个正态总体的均值是否相同,即检验假设

$$H_0: \mu_1=\mu_2=\mu_3=\mu_4$$

一般地,假设因素 A 有 s 个水平 A_1、A_2…,A_s,在水平 $A_j(j=1,2,\cdots,s)$下,进行 n_j($n_j \geqslant 2$)的重复独立试验,得到 n_j 个数据 x_{1j},x_{2j}…,x_{n_jj},(见表 5-13). 假定它们是来自具有相同方差 σ^2,均值为 $\mu_j(j=1,2,\cdots,s)$的正态总体 $N(\mu_j,\sigma^2)$,μ_j,σ^2 未知. 且设不同水平 A_j 下的样本之间相互独立.

表 5-13

水　平	观	察	值			
A_1	x_{11}	x_{21}	…	x_{i1}	…	x_{n_11}
A_2	x_{12}	x_{22}	…	x_{i2}	…	x_{n_22}
⋮	⋮	⋮		⋮		⋮
A_j	x_{1j}	x_{2j}	…	x_{ij}	…	x_{n_jj}
⋮	⋮	⋮		⋮		⋮
A_s	x_{1s}	x_{2s}	…	x_{is}	…	x_{n_ss}

由于 $x_{ij} \sim N(\mu_j, \sigma^2)$，即有 $x_{ij}-\mu_j \sim N(0,\sigma^2)$，故 $x_{ij}-\mu_j$ 可看成是随机误差

记 $x_{ij}-\mu_j=\varepsilon_{ij}$，则 x_{ij} 可写成

$$\begin{cases} x_{ij} = \mu_j + \varepsilon_{ij}, i = 1,2,\cdots,n_j; j = 1,2,\cdots,s \\ \varepsilon_{ij} \sim N(0,\sigma^2) \\ \text{且 } \varepsilon_{ij} \text{ 相互独立}, \mu_j, \sigma^2 \text{ 均为未知参数.} \end{cases} \tag{5.19}$$

(5.19)式称为单因素试验方差分析的数学模型.

判断这个因素对试验指标的影响是否显著就是要检验

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_s$$

方差分析的基本思想就是通过对条件误差和随机误差的比较来判断假设 H_0 是否成立. 如果条件误差比随机误差大得多，我们就认为因素 A 的不同水平对试验结果有显著影响，从而拒绝 H_0；如果条件误差与随机误差的数值差别不大，则认为因素 A 的不同水平对试验结果没有显著影响，就接受 H_0.

为了研究条件误差和随机误差，下面对试验数据的总离差平方和(也称误差平方和)

$$s_T = \sum_{j=1}^{s}\sum_{i=1}^{n_j}(x_{ij} - \bar{x})^2$$

其中

$$\bar{x} = \frac{1}{n}\sum_{j=1}^{s}\sum_{i=1}^{n_j}x_{ij}$$

进行分解，为此令

$$n = \sum_{j=1}^{s}n_j$$

$$\mu = \frac{1}{n}\sum_{j=1}^{s}(n_j\mu_j)$$

$$\bar{x}_{\cdot j} = \frac{1}{n_j}\sum_{i=1}^{n_j}x_{ij}$$

则

$$\begin{aligned} s_T &= \sum_{j=1}^{s}\sum_{i=1}^{n_j}[(x_{ij} - \bar{x}_{\cdot j}) + (\bar{x}_{\cdot j} - \bar{x})]^2 \\ &= \sum_{j=1}^{s}\sum_{i=1}^{n_j}(x_{ij} - \bar{x}_{\cdot j})^2 + 2\sum_{j=1}^{s}\sum_{i=1}^{n_j}(x_{ij} - \bar{x}_{\cdot j})(\bar{x}_{\cdot j} - \bar{x}) + \sum_{j=1}^{s}\sum_{i=1}^{n_j}(\bar{x}_{\cdot j} - \bar{x})^2 \end{aligned}$$

而上式中第二项

$$\begin{aligned} 2\sum_{j=1}^{s}\sum_{i=1}^{n_j}(x_{ij} - \bar{x}_{\cdot j})(\bar{x}_{\cdot j} - \bar{x}) &= 2\sum_{j=1}^{s}(\bar{x}_{\cdot j} - \bar{x})\sum_{i=1}^{n_j}(x_{ij} - \bar{x}_{\cdot j}) \\ &= 2\sum_{j=1}^{s}(\bar{x}_{\cdot j} - \bar{x})\left(\sum_{i=1}^{n_j}x_{ij} - n_j\bar{x}_{\cdot j}\right) = 0 \end{aligned}$$

于是，我们可将 s_T 分解成

$$s_T = s_E + s_A$$

其中
$$s_E = \sum_{j=1}^{s}\sum_{i=1}^{n_j}(x_{ij} - \bar{x}_{.j})^2$$

$$s_A = \sum_{j=1}^{s}\sum_{i=1}^{n_j}(\bar{x}_{.j} - \bar{x})^2 = \sum_{j=1}^{s} n_j(\bar{x}_{.j} - \bar{x})^2$$

s_E 表示在水平 A_j 下，样本值与组内平均值的差异，称为组内离差平方和或组内误差平方和，它是由随机误差引起的；s_A 表示在水平 A_j 下的各组平均值与总平均值的差异，称为组间离差平方和或因素 A 的离差平方和，它是由水平 A_j 及随机误差所引起的. 一般地，若 A 的水平不同引起的条件误差即 μ_i 间的差距显著时，s_A 就会大；反之，当 A 引起的条件误差不显著时，s_A 就会小.

可以证明：(1) s_T 的自由度为 $n-1$，s_E 的自由度为 $n-s$，且 s_T 的自由度可以分解为 s_A 的自由度与 s_E 的自由度之和，即 s_A 的自由度为 $s-1$.

(2) 当 H_0 为真，即 $\mu_1=\mu_2=\cdots=\mu_s$ 时，

$$F = \frac{s_A/(s-1)}{s_E/(n-s)} \sim F(s-1, n-s)$$

其中 $s-1$，$n-s$ 分别是 s_A 和 s_E 的自由度.

若假设 H_0 为真，即 μ_i 间的差距不大时，则因素 A 的各个水平对总体的影响应差不多，s_A 中也应只含有随机误差，因而，F 的值不应太大，如果 F 的值很大，超过临界值 $F_{0.05}$，就应当否定 H_0，即认为 μ_i 之间有显著差异；若 $F>F_{0.01}$时，称为有高度显著差异，或 A 因素高度显著.

上述分析的结果常排成表 5-14 的形式，称为方差分析法.

表 5-14

方差来源	平方和	自由度	均　方	F 值	F_α
因素 A 的影响	s_A	$s-1$	$\bar{s}_A=\frac{s_A}{s-1}$	$F=\frac{\bar{s}_A}{\bar{s}_E}$	$F_\alpha(s-1,n-s)$
误　差	s_E	$n-s$	$\bar{s}_E=\frac{s_E}{n-s}$		
总　和	s_T	$n-1$			

表 5-14 的计算可用 Math cai 软件方便地算出. 下面以本节中的例 1 说明计算方法.

1° 在 DOS 下打入可执行文件名__ VARIANA〈回车〉；

2° 按[F5]建立一个新的数据文件，出现询问：

列数(1—99) ________

输入各水平重复观测次数中的最大值 8，〈回车〉

行数(1—999) ________

输入因子 A 的水平数 4，然后〈回车〉

进入数据编辑窗，开始输入数据，输入数据的格式如下：

	1. x_1	2. x_2	3. x_3	4. x_4	5. x_5	6. x_6	7. x_7	8. x_8
1. 7	1600	1610	1650	1680	1700	1720	1800	
2. 5	1580	1640	1640	1700	1750			
3. 8	1460	1550	1600	1620	1640	1660	1740	1820
4. 6	1510	1520	1530	1570	1600	1680		

数据输入完毕，按[F2]存盘，按[Esc]退出数据编辑窗，返回主菜单；

3° 在主菜单中，按[F9]运行，出现子菜单：

(1)单因子方差分析

(2)无重复观测不考虑交互作用的双因子方差分析

(3)有重复观测考虑交互作用的双因子方差分析

(0)退出

键入 1，2，3 或 0 ________

键入 1，开始方差分析计算，显示结果如下：

《单因子方差分析》

因子 A 的水平数________ 4

水 平	复复观测次数	均 值	平方和
A_1	7	1.6800E+03	2.8600E+04
A_2	5	1.6620E+03	1.6880E+04
A_3	8	1.6362E+02	8.5187E+04
A_4	6	1.5683E+03	2.0683E+04

总均值＝1.6373E+03

方差分析表

来 源	平方和	自由度	均 方	F 值	$F_{0.05}$	$F_{0.01}$
A	4.436 1E+04	3	1.478 7E+04	2.149	3.049	4.817
误 差	1.513 5e+05	22	6.879 6E+03			
总 和	1.957 1E+05	26	7.828 5E+03			

将计算结果存入数据文件？ (Y/N)________

若存入数据文件，键入 Y，并输入数据文件名，否则键入 N.

在方差分析表中的 F 统计量 2.149 小于 F 分布的临界值 $F_{0.05}=3.049$，说明因素 A 作用不显著，应接受假设 $H_0: \mu_1=\mu_2=\mu_3=\mu_4$. 即四种不同的配料方案制成的灯丝对灯泡寿命没有显著影响.

二、无重复不考虑交互作用的双因素试验方差分析

前面介绍了单因素试验方差分析，但实际上在许多实验中常常有几个因素在变化. 例如，在化工生产中，某种原料的用量及反应时间都可能影响产品的产量和质量. 检验药效

时，要用多种剂量的药对不同品种动物试验观察其结果. 因此，需要作多因素试验方差分析.

例 3　为研究如何改善埋在地下的钢管的抗腐蚀性，将两种不同钢管在不同土质中埋了八年，挖出后测得其被腐蚀掉的重量，见表 5-15

表 5-15

		因子 B	
		涂铅的钢管	裸露的钢管
因子 A	细砂土 1	0.18	1.70
	砾砂土 2	0.08	0.21
	淤　泥 3	0.61	1.21
	粘　土 4	0.44	0.89
	沼泽地 5	0.77	0.86
	碱　土 6	1.27	2.64

试问：(1)不同土质对钢管腐蚀的差别是否显著？

(2)不同种类的钢管腐蚀的差别是否显著？

例 4　试验某种钢的不同含铜量在各种温度下的冲击值(单位：公斤・米/厘米^2)，其结果见表 5-16.

表 5-16

含铜量 A / 试验温度 B	0.2%	0.4%	0.8%
20	10.6	11.6	14.5
0	7.0	11.1	13.3
−20	4.2	6.8	11.5
−40	4.2	6.3	8.7

试问不同的试验温度、不同的含铜量下的冲击值是否有显著差异？

这就是两个因子的方差分析问题.

一般地，设因素 A 有 r 个不同水平，$A_1, A_2, \cdots, A_r$，因素 B 有 s 个不同水平 $B_1, B_2, \cdots, B_s$. 我们对两个因素的每对水平组合 (A_i, B_j) 仅进行了一次试验，其试验结果记为 x_{ij} $(i=1,2,\cdots,r; j=1,2,\cdots,s)$，这样共得 rs 个试验结果，见表 5-17

表 5-17

因素 B / 因素 A	B_1	B_2	$\cdots$	B_s
A_1	x_{11}	x_{12}	$\cdots$	x_{1s}
A_2	x_{21}	x_{22}	$\cdots$	x_{2s}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_r	x_{r1}	x_{r2}	$\cdots$	x_{rs}

仍象单因素试验的方差分析一样，假定 $x_{ij} \sim N(\mu_{ij}, \sigma^2)$，即 x_{ij}是从分布为 $N(\mu_{ij}, \sigma^2)$ 的总体中抽得的样本，且各 x_{ij}相互独立，还假定

$$\mu_{ij} = \mu + \alpha_i + \beta_j, i = 1,2,\cdots,r; j = 1,2,\cdots,s$$

这里

$$\mu = \frac{1}{rs}\sum_{i=1}^{r}\sum_{j=1}^{s}\mu_{ij}$$

参数 α_i 表示因素 A 的水平 A_i 的效应，$i=1,2,\cdots,r$；

参数 β_j 表示因素 B 的水平 B_j 的效应，$j=1,2,\cdots,s$，且 α_i, β_j 满足

$$\sum_{i=1}^{r}\alpha_i = 0, \sum_{j=1}^{s}\beta_j = 0 \tag{5.20}$$

则
$$x_{ij} = \mu_{ij} + \varepsilon_{ij}$$

即
$$x_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

由此可得无重复不考虑交互作用的双因素试验方差分析的数学模型为：

$$\begin{cases} x_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \\ \varepsilon_{ij} \sim N(0,\sigma^2), i = 1,2,\cdots,r; j = 1,2,\cdots,s \end{cases} \tag{5.21}$$

这里各 ε_{ij}相互独立，$\mu, \alpha_i, \beta_j, \sigma^2$ 均为未知参数且满足(5.20)式.

要判断因素 A 的影响是否显著就等价于检验假设

$$H_{01}: \alpha_1 = \alpha_2 = \cdots = \alpha_r = 0 \tag{5.22}$$

要判断因子 B 的影响是否显著就等价于检验假设

$$H_{02}: \beta_1 = \beta_2 = \cdots = \beta_s = 0 \tag{5.23}$$

和单因素试验一样，为检验这些假设，我们将试验数据的总离差平方和 s_T 进行分解，令

$$\bar{x}_{i\cdot} = \frac{1}{s}\sum_{j=1}^{s}x_{ij}$$

$$\bar{x}_{\cdot j} = \frac{1}{r}\sum_{i=1}^{r}x_{ij}$$

$$\bar{x} = \frac{1}{rs}\sum_{i=1}^{r}\sum_{j=1}^{s}x_{ij} = \frac{1}{r}\sum_{i=1}^{r}\bar{x}_{i\cdot} = \frac{1}{s}\sum_{j=1}^{s}\bar{x}_{\cdot j}$$

则

$$\begin{aligned} s_T &= \sum_{i=1}^{r}\sum_{j=1}^{s}(x_{ij} - \bar{x})^2 \\ &= \sum_{i=1}^{r}\sum_{j=1}^{s}[(x_{ij} - \bar{x}_{i\cdot} - \bar{x}_{\cdot j} + \bar{x}) + (\bar{x}_{i\cdot} - \bar{x}) + (\bar{x}_{\cdot j} - \bar{x})]^2 \end{aligned}$$

$$= \sum_{i=1}^{r}\sum_{j=1}^{s}(x_{ij} - \bar{x}_{i\cdot} - \bar{x}_{\cdot j} + \bar{x})^2 + s\sum_{i=1}^{r}(\bar{x}_{i\cdot} - \bar{x})^2 + r\sum_{j=1}^{s}(\bar{x}_{\cdot j} - \bar{x})^2$$
$$= s_E + s_A + s_B$$

其中
$$s_E = \sum_{i=1}^{r}\sum_{j=1}^{s}(x_{ij} - \bar{x}_{i\cdot} - \bar{x}_{\cdot j} + \bar{x})^2$$

$$s_A = s\sum_{i=1}^{r}(\bar{x}_{i\cdot} - \bar{x})^2$$

$$s_B = r\sum_{j=1}^{s}(\bar{x}_{i\cdot} - \bar{x})^2$$

上式推导中,任意两个交叉乘积项对 i,j 取和的结果均为零.

s_E 称为误差平方和,它反映了各种随机因素所引起的误差. s_A 称为因素 A 的离差平方和,它反映了因素 A 的不同水平对试验结果的影响. s_B 称为因素 B 的离差平方和,它反映了因素 B 的不同水平对试验结果的影响.

可以证明:(1) s_T 的自由度为 $rs-1$, s_E 的自由度为 $(r-1)(s-1)$, s_A 的自由度为 $r-1$, s_B 的自由度为 $s-1$.

(2) 当 H_{01} 为真,即 $\alpha_1=\alpha_2=\cdots=\alpha_r=0$ 为真时

$$F_A = \frac{s_A/(r-1)}{s_E/[(r-1)(s-1)]} \sim F(r-1,(r-1)(s-1))$$

当 H_{02} 为真,即 $\beta_1=\beta_2=\cdots=\beta_s=0$ 为真时

$$F_B = \frac{s_B/(s-1)}{s_E/[(r-1)(s-1)]} \sim F(s-1,(r-1)(s-1))$$

于是,取显著性水平为 α,得假设 H_{01} 的拒绝域为

$$F_A = \frac{s_A/(r-1)}{s_E/[(r-1)(s-1)]} \geqslant F_\alpha(r-1,(r-1)(s-1))$$

假设 H_{02} 的拒绝域为

$$F_B = \frac{s_B/(s-1)}{s_E/[(r-1)(s-1)]} \geqslant F_\alpha(s-1,(r-1)(s-1))$$

其方差分析表如表 5-18 所示.

表 5-18

方差来源	平方和	自由度	均　方	F 值	F_α
因素 A 的影响	s_A	$r-1$	$\bar{s}_A=\frac{s_A}{r-1}$	$F_A=\frac{\bar{s}_A}{\bar{s}_E}$	$F_\alpha(r-1,(r-1)(s-1))$
因素 B 的影响	s_B	$s-1$	$\bar{s}_B=\frac{s_B}{s-1}$	$F_B=\frac{\bar{s}_B}{\bar{s}_E}$	$F_\alpha(r-1,(r-1)(s-1))$
误　差	s_E	$(r-1)(s-1)$	$\bar{s}_E=\frac{s_E}{(r-1)(s-1)}$		
总　和	s_T	$rs-1$			

下面以例 3 说明用 Math cai 软件的计算步骤.

1°、2° 与单因素方差分析的计算步骤相同，只不过在出现询问列数时，输入因素 B 的水平数 2；输入数据的格式如下：

.	1. x_1	2. x_2
1.	0.18	1.70
2.	0.08	0.21
3.	0.61	1.21
4.	0.44	0.89
5.	0.77	0.86
6.	1.27	2.64

在 3°中，键入 2，开始方差分析计算，显示结果如下：

《无重复观测的双因子方差分析》

因子 A 的水平数________________6

因子 B 的水平数________________2

水 平	均 值	平方和	水 平	均 值	平方和
A_1	9.400 0E−01	1.155 2E+00	B_1	5.583 3E−01	9.398 8E−01
A_2	1.450 0E−01	8.450 0E−03	B_2	1.251 7E+00	3.499 5E+00
A_3	9.100 0E−01	1.800 0E−01			
A_4	6.650 0E−01	1.012 5E−01			
A_5	8.150 0E−01	4.050 0E−03			
A_6	1.955 0E+00	9.384 5E−01			

总均值＝9.050 0E−01

按任意键继续……

方差分析表

来 源	平方和	自由度	均 方	F 值	$F_{0.05}$	$F_{0.01}$
A	3.494 1E+00	5	6.988 2E−01	3.696	5.050	10.967
B	1.442 1E+00	1	1.442 1E+00	7.628	6.608	16.258
误 差	9.452 7E−01	5	1.890 5E−01			
总 和	5.881 5E+00	11	5.346 8E−01			

将计算结果存入数据文件？ (Y/N)______

由方差分析表易见，$F_A=3.696$，$F_{0.05}(5,5)=5.050$，此时 $F_A<F_{0.05}(5,5)$，说明因素 A，即不同土质对钢管腐蚀的差别是不显著的；$F_B=7.628$，$F_{0.05}(1,5)=6.608$，此时 $F_A>F_{0.05}(1,5)$，说明因素 B，即不同种类的钢管腐蚀的差别是显著的.

三、有重复考虑交互作用的双因素试验方差分析

模型(5.21)的优点是不需做重复试验，即对 A,B 每一水平组合只需做一次试验，这大大节省了人力、物力，但也有缺点.因在实际问题中，不仅因素对试验结果有影响，而且因素之间还会联合起来对结果产生新的影响，即因素之间有交互作用.如下例所示.

例 5 一火箭使用了四种燃料、三种推进器做射程试验.每种燃料与每种推进器的组合各发射火箭两次，得结果如表 5-19 所示(射程以海里计).

表 5-19

燃　料	推进器					
	B_1		B_2		B_3	
A_1	58.2,	52.6	56.2,	41.2	65.3,	60.8
A_2	49.1,	42.8	54.1,	50.5	51.6,	48.4
A_3	60.1,	58.3	70.9,	73.2	39.2,	40.7
A_4	75.8,	71.5	58.2,	51.0	48.7,	41.4

试分析推进器和燃料这两个因素对射程是否有显著的影响？交互作用是否显著？

若以 A_i 与 B_j 组合两次试验的平均值为 x_{ij}，按模型(5.21)分析计算，可得方差分析表如下：

方差分析表

来　源	平方和	自由度	均　方	F 值	$F_{0.05}$	$F_{0.01}$
A	1.308 4E+02	3	4.361 3E+01	0.296	4.757	9.780
B	1.854 9E+02	2	9.274 5E+01	0.629	5.143	10.925
误　差	8.843 5E+02	6	1.473 9E+02			
总　和	1.200 7E+03	11	1.091 5E+02			

由方差分析表中结果易见，推进器和燃料对射程的影响不显著，也就是说在不同水平下的推进器和不同水平下的燃料对射程的影响不大，但从表 5-19 中试验结果可见，B_1 和 A_4 的组合射程很远，B_3 和 A_3 的组合射程很近.之所以 F 检验得出的结论即推进器和燃料对射程的影响不显著，是由于不同水平的值作了平均的原故.为了反映两种因素搭配后的影响，必须引入“交互作用”的概念.

表 5-20

因素 B / 因素 A	B_1	B_2	…	B_s
A_1	$x_{111},x_{112},\cdots,x_{11t}$	$x_{121},x_{122},\cdots,x_{12t}$	…	$x_{1s1},x_{1s2},\cdots,x_{1st}$
A_2	$x_{211},x_{212},\cdots,x_{21t}$	$x_{221},x_{222},\cdots,x_{22t}$	…	$x_{2s1},x_{2s2},\cdots,x_{2st}$
…	…	…	…	…
A_r	$x_{r11},x_{r12},\cdots,x_{r1t}$	$x_{r21},x_{r22},\cdots,x_{r2t}$	…	$x_{rs1},x_{rs2},\cdots,x_{rst}$

考察 A、B 两个因素之间是否存在交互作用，只凭一次试验是不行的，至少需要重复两次以上，方能分析有无交互作用．这种分析对选择最优生产方案有重大作用．为了便于说明这种分析方法，假设对两个因素的每对水平组合 $(A_i,B_j)(i=1,2,\cdots,r;j=1,2,\cdots,s)$ 都作 $t(t\geqslant 2)$ 次试验，（称为等重复试验）．这样共得到 rst 个试验结果，见表 5-20．

其中 x_{ijk} 表示对因素 A 的第 i 个水平 A_i，因素 B 的第 j 个水平 B_j 的第 k 次试验结果．

假定：

$x_{ijk}\sim N(\mu_{ij},\sigma^2)$，$i=1,2,\cdots,r;j=1,2,\cdots,s;k=1,2,\cdots,t$．且各 x_{ijk} 独立．

这里 μ_{ij}，σ^2 均为未知参数．或写成

$$\begin{cases} x_{ijk}=\mu_{ij}+\varepsilon_{ijk}, i=1,2,\cdots,r;j=1,2,\cdots,s \\ \varepsilon_{ijk}\sim N(0,\sigma^2) \quad k=1,2,\cdots,t \\ \text{各 } x_{ijk} \text{ 独立} \end{cases} \tag{5.24}$$

记 $\mu=\frac{1}{rs}\sum_{i=1}^{r}\sum_{j=1}^{s}\mu_{ij}$

$\mu_{i\cdot}=\frac{1}{s}\sum_{j=1}^{s}\mu_{ij}, i=1,2,\cdots,r$

$\mu_{\cdot j}=\frac{1}{r}\sum_{i=1}^{r}\mu_{ij}, j=1,2,\cdots,s$

$\alpha_i=\mu_{i\cdot}-\mu, i=1,2,\cdots,r$

$\beta_j=\mu_{\cdot j}-\mu, j=1,2,\cdots,s$

易见

$$\sum_{i=1}^{r}\alpha_i=0,\quad \sum_{j=1}^{s}\beta_j=0$$

称 μ 为总平均，称 α_i 为水平 A_i 的效应，称 β_j 为水平 B_j 的效应．这样可将 μ_{ij} 表示成

$$\mu_{ij}=\mu+\alpha_i+\beta_j+(\mu_{ij}-\mu_{i\cdot}-\mu_{\cdot j}+\mu) \quad i=1,2,\cdots,r;j=1,2,\cdots,s$$

记 $\delta_{ij}=\mu_{ij}-\mu_{i\cdot}-\mu_{\cdot j}+\mu$，则

$$\mu_{ij}=\mu+\alpha_i+\beta_j+\delta_{ij} \quad i=1,2,\cdots,r;j=1,2,\cdots,s$$

δ_{ij} 称为水平 A_i 和水平 B_j 的交互效应，这是由 A_i，B_j 搭配起来联合起作用而引起的．易见

$$\sum_{i=1}^{r}\delta_{ij}=0, j=1,2,\cdots,s$$

$$\sum_{j=1}^{s}\delta_{ij}=0, i=1,2,\cdots,r$$

通过以上分析，有重复考虑交互作用的双因素试验方差分析的数学模型可写为：

$$\begin{cases} x_{ijk}=\mu+\alpha_i+\beta_j+\delta_{ij}+\varepsilon_{ijk} \\ \varepsilon_{ijk}\sim N(0,\sigma^2), i=1,2,\cdots,r;j=1,2,\cdots,s \\ \qquad\qquad k=1,2,\cdots,t \\ \text{各 } \varepsilon_{ijk} \text{ 独立} \\ \sum_{i=1}^{r}\alpha_i=0, \sum_{j=1}^{s}\beta_j=0, \sum_{i=1}^{r}\delta_{ij}=0, \sum_{j=1}^{s}\delta_{ij}=0 \end{cases} \tag{5.25}$$

其中 $\mu,\alpha_i,\beta_j,\delta_{ij},\sigma^2$ 均为未知参数.

对于这一模型，要判断因素 A、B 及交互作用 $A\times B$ 的影响是否显著，分别等价于检验假设

$$H_{01}:\alpha_1=\alpha_2=\cdots=\alpha_r=0$$
$$H_{02}:\beta_1=\beta_2=\cdots=\beta_s=0$$
$$H_{03}:\delta_{ij}=0, \text{对一切 } i=1,2,\cdots,r; j=1,2,\cdots,s$$

为了检验这些假设，将试验数据的总离差平方和 s_T 进行分解，令

$$\bar{x}=\frac{1}{rst}\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}x_{ijk}$$

$$\bar{x}_{ij\cdot}=\frac{1}{t}\sum_{k=1}^{t}x_{ijk}, i=1,2,\cdots,r; j=1,2,\cdots,s$$

$$\bar{x}_{i\cdot\cdot}=\frac{1}{st}\sum_{j=1}^{s}\sum_{k=1}^{t}x_{ijk}, i=1,2,\cdots,r$$

$$\bar{x}_{\cdot j\cdot}=\frac{1}{rt}\sum_{i=1}^{r}\sum_{k=1}^{t}x_{ijk}, j=1,2,\cdots,s$$

于是，总离差平方和可以分解为：

$$\begin{aligned}s_T&=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}(x_{ijk}-\bar{x})^2\\&=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}[(x_{ijk}-\bar{x}_{ij\cdot})+(\bar{x}_{i\cdot\cdot}-\bar{x})\\&+(\bar{x}_{\cdot j\cdot}-\bar{x})+(\bar{x}_{ij\cdot}-\bar{x}_{i\cdot\cdot}-\bar{x}_{\cdot j\cdot}+\bar{x})]^2\\&=s_E+s_A+s_B+s_{A\times B}\end{aligned}$$

其中

$$s_E=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}(x_{ijk}-\bar{x}_{ij\cdot})^2$$

$$s_A=st\sum_{i=1}^{r}(\bar{x}_{i\cdot\cdot}-\bar{x})^2$$

$$s_B=rt\sum_{j=1}^{s}(\bar{x}_{\cdot j\cdot}-\bar{x})^2$$

$$s_{A\times B}=t\sum_{i=1}^{r}\sum_{j=1}^{s}(\bar{x}_{ij\cdot}-\bar{x}_{i\cdot\cdot}-\bar{x}_{\cdot j\cdot}+\bar{x})^2$$

s_E 称为误差平方和，s_A、s_B、$s_{A\times B}$分别是因素 A、B、A 与 B 交互作用的离差平方和.

可以证明：(1) s_T 的自由度为 $rst-1$，s_E 的自由度为 $rs(t-1)$，s_A 的自由度为 $r-1$，s_B 的自由度为 $s-1$，$s_{A\times B}$的自由度为$(r-1)(s-1)$.

(2) 当 H_{01} 为真，即 $\alpha_1=\alpha_2=\cdots=\alpha_r=0$ 为真时，

$$F_A=\frac{s_A/(r-1)}{s_E/(rs(t-1))}\sim F(r-1,rs(t-1))$$

取显著性水平为 α 时，得假设 H_{01}的拒绝域为

$$F_A\geqslant F_\alpha(r-1,rs(t-1))$$

当 H_{02}为真，即 $\beta_1=\beta_2=\cdots=\beta_s=0$ 为真时，

$$F_B = \frac{s_B/(s-1)}{s_E/(rs(t-1))} \sim F(s-1, rs(t-1))$$

类似地，在显著性水平 α 下，假设 H_{02} 的拒绝域为

$$F_B \geqslant F_\alpha(s-1, rs(t-1))$$

当 H_{03} 为真，即 $\delta_{ij}=0, i=1,2,\cdots,r; j=1,2,\cdots,s$ 为真时

$$F_{A\times B} = \frac{s_{A\times B}/((r-1)(s-1))}{s_E/(rs(t-1))} \sim F((r-1)(s-1), rs(t-1))$$

在显著性水平 α 下，假设 H_{03} 的拒绝域为

$$F_{A\times B} \geqslant F_\alpha((r-1)(s-1), rs(t-1))$$

上述结果可汇总成方差分析表 5-21

表 5-21

方差来源	平方和	自由度	均　方	F 值	F_α
因素 A	s_A	$r-1$	$\bar{s}_A=\frac{s_A}{r-1}$	$F_A=\frac{\bar{s}_A}{\bar{s}_E}$	$F_\alpha(r-1, rs(t-1))$
因素 B	s_B	$s-1$	$\bar{s}_B=\frac{s_B}{s-1}$	$F_B=\frac{\bar{s}_B}{\bar{s}_E}$	$F_\alpha(s-1, rs(t-1))$
$A\times B$	$s_{A\times B}$	$(r-1)(s-1)$	$\bar{s}_{A\times B}=\frac{s_{A\times B}}{(r-1)(s-1)}$	$F_{A\times B}=\frac{\bar{s}_{A\times B}}{\bar{s}_E}$	$F_\alpha((r-1)(s-1), rs(t-1))$
误　差	s_E	$rs(t-1)$	$\bar{s}_E=\frac{s_E}{rs(t-1)}$		
总　和	s_T	$rst-1$			

下面以例 5 说明用 Math cai 软件的计算步骤.

1°,2°与单因素方差分析的计算步骤相同，只不过在出现询问列数时，输入因素 B 的水平数 $s=3$ 与重复观测次数 $t=2$ 的乘积 6.

输入数据格式如下：

2	1.	2.	3.	4.	5.	6.
1.	58.2	52.6	56.2	41.2	65.3	60.8
2.	49.1	42.8	54.1	50.5	51.6	48.4
3.	60.1	58.3	70.9	73.2	39.2	40.7
4.	75.8	71.5	58.2	51.0	48.7	41.4

在 3°中，键入 3，开始方差分析计算，显示结果如下：

《有重复观测的双因子方差分析》

因子 A 的水平数＿＿＿＿＿＿＿＿ 4

因子 B 的水平数＿＿＿＿＿＿＿＿ 3

重复观测次数＿＿＿＿＿＿＿＿ 2

水 平	均 值	平方和	水 平	均 值	平方和
A_1	5.571 7E+01	2.062 2E+02	B_1	5.855 0E+01	7.942 3E+02
A_2	4.941 7E+01	4.134 3E+01	B_2	5.691 3E+01	6.464 2E+02
A_3	5.706 7E+01	1.044 1E+03	B_3	4.951 3E+01	5.897 1E+02
A_4	5.776 7E+01	8.480 4E+02			

总均值=5.499 2E+01

方差分析表

来 源	平方和	自由度	均 方	F 值	$F_{0.05}$	$F_{0.01}$
A	2.616 7E+02	3	8.722 5E+01	4.417	3.490	5.953 *
B	3.709 8E+02	2	1.854 9E+02	9.394	3.885	6.927 * *
A * B	1.768 7E+03	6	2.947 8E+02	14.929	2.996	4.821 * *
误 差	2.369 5E+02	12	1.974 6E+01			
总 和	2.638 3E+03	23	1.147 1E+02			

将计算结果存入数据文件?　　　　　　　　　　(Y/N)______

由方差分析表易见,$F_A=4.417>F_{0.05}(3,12)=3.490$,但 $F_A=4.417<F_{0.01}(3,12)=5.953$,说明燃料对射程的影响显著;$F_B=9.394>F_{0.05}(3,12)=3.885$,且 $F_B>F_{0.01}(3,12)=6.927$,说明推进器对射程的影响是高度显著的;$F_{A\times B}=14.929>F_{0.05}(6,12)=2.996$,且 $F_{A\times B}>F_{0.01}(6,12)=4.821$,说明交互作用对射程的影响是高度显著的.

习题五

1. 某厂有一种新产品,其推销策略有 s_1,s_2,s_3 三种可供选择,但方案所需资金、时间都不同,加上市场情况的差别,因而获利和亏损情况不同,而市场情况也有三种:N_1 需求量大,N_2 需求量一般,N_3 需求量低.市场情况的概率并不知道,其效益值见表 5-22.

表 5-22

效益值 \ 状态 / 方案		市场情况		
		N_1	N_2	N_3
销售策略	s_1	50	10	−5
	s_2	30	25	0
	s_3	10	10	10

(1)用乐观法进行决策;

(2)用悲观法进行决策;

(3)用等可能法进行决策;

2. 某工程采用正常速度施工,若无坏天气的影响,可确保在 30*d*(*d* 为时间单位:天)内按期完成工程,但据天气预报,15*d* 后天气肯定变坏,有 40%的可能出现阴雨天气,但这不会影响工程进度;有 50%的可能遇到小风暴而使工期推迟 15*d*;另有 10%的可能遇到大风暴而使工期推迟 20*d*.关于以上可能出现的情况,考虑两种方案:

(1)提前加班,确定工程在 15*d* 内完成,实施此方案需增加额外支付 18 000 元;

(2)先维持原定的施工进度,等 15*d* 后根据实际出现的天气状况再作决策;

1)若遇阴雨天,则维持正常进度,不必支付额外费用.

2)若遇小风暴,则有下述两个供选方案:一是抽空(风暴过后)施工,支付工程延期损失费 20 000 元,二是采取应急措施,实施此措施可能有三种结果:有 50%的可能减少误工期 1d,支付延期损失费和应急费用 24 000 元,有 30%的可能减少误工期 2d,支付延期损失费和应急费用 18 000 元;有 20%的可能减少误工期 3d,支付延期损失费和应急费用 12 000 元.

3)若遇大风暴,则仍有两个方案可供选择:一是抽空进行施工,支付工程延期损失费 50 000 元;二是采取应急措施,实施此措施可能有三种结果:有 70%的可能减少误工期 2d,支付延期损失费及应急费用 54 000 元,有 20%的可能减少误工期 3d,支付延期损失费及应急费 46 000 元;有 10%的可能减少误工期 4d,支付延期损失费及应急费 38 000 元.

试进行决策,选择最佳行动方案.

3. 铝合金化学铣切工艺中,为了便于生产操作,需要对腐蚀速度进行控制,因此要考虑腐蚀液温度(℃),碱浓度(克/升),腐蚀液含铝量(克/升)对腐蚀速度的影响.一共做了 44 次试验,所得如表 5-23 所示.

试研究腐蚀液温度、碱浓度、含铝量与腐蚀速度之间的关系.

表 5-23

试验号	温度(x_1)	含铝量(x_2)	碱浓度(x_3)	腐蚀速度(y)
1	73	12	200	0.024 0
2	73	21	200	0.023 5
3	75	30	200	0.024 0
4	75	36	200	0.024 5
5	75	42	200	0.019
6	75	48	200	0.018 5
7	79	12	200	0.032
8	79	21	200	0.030
9	79	30	200	0.029
10	79	36	200	0.027 5
11	79	42	200	0.025
12	79	48	200	0.022 5
13	83	12	200	0.037 0
14	83	21	200	0.036
15	83	30	200	0.035 5
16	83	36	200	0.032 5
17	83	42	200	0.030 5
18	83	48	200	0.027 0
19	87	12	200	0.044 0
20	87	21	200	0.042 5
21	87	30	200	0.042 0
22	87	36	200	0.039 0
23	87	42	200	0.036 0
24	87	46	200	0.032 5
25	77	19	150	0.023 0
26	77	19	175	0.025 0
27	77	19	200	0.026 5

试验号	温度(x_1)	含铝量(x_2)	碱浓度(x_3)	腐蚀速度(y)
28	77	19	225	0.028 5
29	77	19	250	0.029 0
30	81	27	150	0.028 5
31	81	27	175	0.029 5
32	81	27	200	0.031 0
33	81	27	225	0.031 5
34	81	27	250	0.032 0
35	85	35	150	0.034 5
36	85	35	176	0.035 5
37	85	35	200	0.037 0
38	85	35	225	0.039 0
39	85	35	260	0.040 5
40	89	43	150	0.037 5
41	89	43	175	0.038 0
42	89	43	200	0.040 0
43	89	43	225	0.043 0
44	89	43	250	0.045 0

4. 经钻探某地区煤矿上表面高度数据如表 5-24. 其中 x 为横坐标，y 为纵坐标，为了作趋势面分析，试研究上表面高度 h 与横坐标 x，纵坐标 y 的关系.

表 5-24

序　号	x	y	h
1	2.00	6.0	−11
2	2.00	5.5	0
3	2.00	5.0	10
4	2.00	4.5	20
5	2.00	4.0	31
6	2.25	6.0	−11
7	2.25	5.5	−1
8	2.25	5.0	9
9	2.25	4.5	19
10	2.25	4.0	30
11	2.50	6.0	−12
12	2.50	5.5	−2
13	2.50	5.0	8
14	2.50	4.50	19
15	2.50	4.0	29
16	2.75	6.0	−13
17	2.75	5.5	−2
18	2.75	5.0	7
19	2.75	4.5	18
20	2.75	4.0	28
21	3.00	6.0	−14
22	3.00	5.5	13
23	3.00	5.0	6
24	3.00	4.5	17
25	3.00	4.0	27

5. 生活在不同地区居民的血液中胆固醇的含量一般是不相同的. 表 5-25 给出了生活在三个不同地区居民的血液中胆固醇的含量. 试检验这三个地区居民血液中胆固醇的平均量之间是否有显著差别?

表 5-25

地　区	测	量	值							
1	403	311	269	336	259					
2	312	222	302	420	420	386	353	210	286	290
3	403	244	353	235	319	260				

6. 某种化工过程在三种浓度, 四种温度水平下得率的数据如表 5-26 所示. 试在水平 $\alpha=0.05$ 下, 检验不同浓度, 不同温度下的得率是否有显著的差异? 交互作用是否显著?

表 5-26

浓度(%) \ 温度	10℃		24℃		38℃		52℃	
2	14	10	11	11	13	9	10	12
4	9	7	10	8	7	11	6	10
6	5	11	13	14	12	13	14	10

附　录　1992～1997 年全国大学生数学模型竞赛题

1992 年全国大学生数学模型联赛赛题

A 题　施肥效果分析

某地区作物生长所需的营养素主要是氮(N)、磷(P)、钾(K)。某作物研究所在该地区对土豆与生菜做了一定数量的实验，实验数据如下列表格所示，其中 ha 表示公顷，t 表示吨，kg 表示公斤。当一个营养素的施肥量变化时，总将另二个营养素的施肥量保持在第七个水平上，如对土豆产量关于 N 的施肥量做实验时，P 与 K 的施肥量分别取为 196kg/ha 与 372kg/ha。

试分析施肥量与产量之间的关系，并对所得结果从应用价值与如何改进等方面作出估价。

土豆：

N

施肥量 (kg/ha)	产量 (t/ha)
0	15.18
34	21.36
67	25.72
101	32.29
135	34.03
202	39.45
259	43.15
336	43.46
404	40.83
471	30.75

P

施肥量 (kg/ha)	产量 (t/ha)
0	33.46
24	32.47
49	36.06
73	37.96
98	41.04
147	40.09
196	41.26
245	42.17
294	40.36
342	42.73

K

施肥量 (kg/ha)	产量 (t/ha)
0	18.98
47	27.35
93	34.86
140	38.52
186	38.44
279	37.73
372	38.43
465	43.87
558	42.77
651	46.22

生菜：

N

施肥量 (kg/ha)	产量 (t/ha)
0	11.62
28	12.70
56	14.56
84	16.27
112	17.25
168	22.59
224	21.63
280	19.34
336	16.12
392	14.11

P

施肥量 (kg/ha)	产量 (t/ha)
0	6.39
49	9.48
98	12.46
147	14.33
195	17.10
294	21.94
391	22.64
489	21.34
587	22.07
685	24.53

K

施肥量 (kg/ha)	产量 (t/ha)
0	15.75
47	16.76
93	16.89
140	16.24
186	17.56
279	19.20
372	17.97
465	15.84
558	20.11
651	19.40

B 题　实验数据分解

组成生命蛋白质的若干种氨基酸可以形成不同的组合。通过质谱实验测定分子量来分析某个生命蛋白质分子的组成时，遇到的首要问题就是如何将它的分子量 X 分解为几个氨基酸的已知分子量 $a[i](i=1,2,\cdots,n)$ 之和。某实验室所研究的问题中：

$n=18$

$a[1:18]$ = 57,71,87,97,99,101,103,113,114,115,128,129,131,137,147,156,163,186。

X 为正整数 $\leqslant 1\ 000$

要求针对该实验室拥有或不拥有微型计算机的情况，对上述问题提出你们的解答，并就你所研讨的数学模型与方法在一般情形下进行讨论。

1993 年全国大学生数学模型竞赛试题

A 题　非线性交调的频率设计

如果一非线性器件的输入 $u(t)$ 与输出 $y(t)$ 的关系是 $y(t)=u(t)+u^2(t)$（其中 t 是时间），那么当输入是包含频率 f_1,f_2 的信号 $u(t)=\cos 2\pi f_1 t+\cos 2\pi f_2 t$ 时，输出 $y(t)$ 中不仅包含输入信号 f_1,f_2，而且还会出现 $2f_1,f_2\pm f_2$ 等新的频率成分，这些新的频率称为交调，如果交调出现在原有频率 f_1,f_2 的附近，就会形成噪声干扰，因此工程设计中对交调的出现有一定的要求。

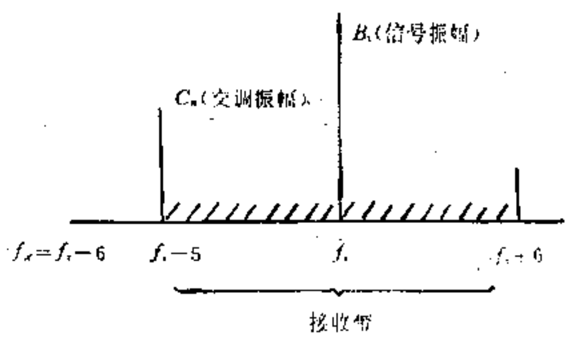
现有一 SCS(非线性) 系统，其输入输出关系由如下一组数据给出：

输入 u	0	5	10	20	30	40	50	60	80
输出 y	0	2.25	6.80	20.15	35.70	56.40	75.10	87.85	98.50

输入信号为 $u(t)=A_1\cos 2\pi f_1 t+A_2\cos 2\pi f_2 t+A_3\cos 2\pi f_3 t$，其中 $A_1=25,A_2=10,A_3=45$ 是输入信号振幅，对输入信号的频率 f_1,f_2,f_3 的设计要求为：

1) $36\leqslant f_1\leqslant 40, 41\leqslant f_2\leqslant 50, 46\leqslant f_3\leqslant 53$。

2) 输出中的交调均不得出现在 $f_i\pm 5$ 的范围内 $(i=1,2,3)$，此范围称为 f_i 的接收带（参见下图）。



3) 定义输出中的信噪比 $SNR = 10\log_{10}\frac{B_i^2}{C_n^2}$(单位:分贝),其中 B_i 是输出中对应于频率为 f_i 的信号的振幅,C_n 是某一频率为 f_n 的交调的振幅,若 f_n 出现在 $f_n = f_i \pm 6$ 处($i = 1,2,3$),则对应的 SNR 应大于 10 分贝(参见上图)。

4) f_i 不得出现在 f_j 的接收带内($i,j = 1,2,3, i \neq j$)。

5) 为简单起见,f_i 只取整数值,且交调只考虑 2 阶类型(即 $\{f_i \pm f_j\}, i,j = 1,2,3$) 和 3 阶类型(即 $\{f_i \pm f_j \pm f_k\}, i,j,k = 1,2,3$)。

试按上述要求设计输入信号频率 f_1, f_2, f_3。

B 题 足球队排名次

下表给出了我国 12 支足球队在 1988～1989 年全国足球甲级联赛中的成绩,要求

1) 设计一个依据这些成绩排出诸队名次的算法,并给出用该算法排名次的结果。

2) 把算法推广到任意 N 个队的情况。

3) 讨论:数据应具有什么样的条件,用你的方法才能够排出诸队的名次。

对下表的说明:

1)12 支球队依次记为 $T_1, T_2, \cdots, T_{12}$。

2) 符号 X 表示两队未曾比赛。

3) 数字表示两队比赛结果,如 T_3 行与 T_8 列交叉处的数字表示:T_3 与 T_8 比赛了 2 场;T_3 与 T_8 的进球数之比为 0:1 和 3:1。

	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}	T_{11}	T_{12}
T_1	X	0:1 1:0 0:0	2:2 1:0 0:2	2:0 3:1 1:0	3:1	1:0	0:1 1:3	0:2 2:1	1:0 4:0	1:1 1:1	X	X
T_2		X	2:0 0:1 1:3	0:0 2:0 0:0	1:1	2:1	1:1 1:1	0:0 0:0	2:0 1:1	0:2 0:0	X	X
T_3			X	4:2 1:1 0:0	2:1	3:0	1:0 1:4	0:1 3:1	1:0 2:3	0:1 2:0	X	X
T_4				X	2:3	0:1	0:5 2:3	2:1 1:3	0:1 0:0	0:1 1:1	X	X
T_5					X	0:1	X	X	X	X	1:0 1:2	0:1 1:1
T_6						X	X	X	X	X	X	
T_7							X	1:0 2:0 0:0	2:1 3:0 1:0	3:1 3:0 2:2	3:1	2:0
T_8								X	0:1 1:2 2:0	1:1 1:0 0:1	3:1	2:0
T_9										3:0 1:0 0:0	1:0	1:0
T_{10}											1:0	2:0
T_{11}											X	1:1 1:2
T_{12}												1:1 X

1994 年全国大学生数学建模竞赛

A 题　逢山开路

要在一山区修建公路，首先测得一些地点的高程，数据见表 1（平面区域 $0 \leqslant x \leqslant 5\,600, 0 \leqslant y \leqslant 4\,800$，表中数据为坐标点的高程，单位：米）。数据显示：在 $y = 3\,200$ 处有一东西走向的山峰；从坐标（2 400，2 400）到（4 800，0）有一西北——东南走向的山谷；在（2 000，2 800）附近有一山口湖，其最高水位略高于 1 350 米，雨季在山谷中形成一溪流。经调查知，雨量最大时溪流水面宽度 W 与（溪流最深处的）x 坐标的关系可近似表示为

$$w(x) = \left(\frac{x - 2\,400}{2}\right)^{3/4} + 5 \quad (2\,400 \leqslant x \leqslant 4\,000)$$

公路从山脚（0，800）处开始，经居民点（4 000，2 000）至矿区（2 000，4 000）。已知路段公程成本及对路段坡度 α（上升高程与水平距离之比）的限制如表 2。

1）试给出一种线路设计方案，包括原理、方法及比较精确的线路位置（含桥梁、隧道），并估算该方案的总成本。

2）如果居民点改为 $3600 \leqslant x \leqslant 4000, 2000 \leqslant y \leqslant 2400$ 的居民区，公路只段经过居民区即可，那么你的方案有什么改变。

表 1　↑北

Y/X	0	400	800	1 200	1 600	2 000	2 400	2 800	3 200	3 600	4 000	4 400	4 800	5 200	5 600
4 800	1 350	1 370	1 390	1 400	1 410	960	940	880	800	690	570	430	290	210	150
4 400	1 370	1 390	1 410	1 430	1 440	1 140	1 110	1 050	950	820	690	540	380	300	210
4 000	1 380	1 410	1 430	1 450	1 470	1 320	1 280	1 200	1 080	940	780	620	460	370	350
3 600	1 420	1 430	1 450	1 480	1 500	1 550	1 510	1 430	1 300	1 200	980	850	750	550	500
3 200	1 430	1 450	1 460	1 500	1 550	1 600	1 550	1 600	1 600	1 600	1 500	1 500	1 500	1 550	1 500
2 800	950	1 190	1 370	1 500	1 200	1 100	1 550	1 600	1 550	1 380	1 070	900	1 050	1 150	1 200
2 400	910	1 090	1 270	1 500	1 200	1 100	1 350	1 450	1 200	1 150	1 010	880	1 000	1 050	1 100
2 000	880	1 060	1 230	1 390	1 500	1 500	1 400	900	1 100	1 060	950	870	900	930	950
1 600	830	980	1 180	1 320	1 450	1 420	1 400	1 300	700	900	850	840	380	780	750
1 200	740	880	1 080	1 130	1 250	1 280	1 230	1 040	900	500	700	780	750	650	550
800	650	760	880	970	1 020	1 050	1 020	830	800	700	300	500	550	480	350
400	510	620	730	800	850	870	850	780	720	650	500	200	300	350	320
0	370	470	550	600	670	690	670	620	580	450	400	300	100	150	250

表 2

工程种类	一般路段	桥梁	隧道
工程成本（元 / 米）	300	2 000	1 500（长度 $\leqslant$ 300 米）；3 000（长度 $>$ 300 米）。
对坡度 α 的限制	$\alpha < 0.125$	$\alpha = 0$	$\alpha < 0.100$

B题　锁具装箱

某厂生产一种弹子锁具，每个锁具的钥匙有5个槽，每个槽的高度从{1,2,3,4,5,6}6个数(单位略)中任取一数。由于工艺及其它原因，制造锁具时对5个槽的高度还有两个限制：至少有3个不同的数；相邻两槽的高度之差不能为5。满足以上条件造出来的所有互不相同的锁具称为一批。

从顾客的利益出发，自然希望在每批锁具中"一把钥匙开一把锁"。但是在当前工艺条件下，对于同一批中两个锁具是否能够互开，有以下试验结果：若二者相对应的5个槽的高度中有4个相同，另一个槽的高度差为1，则可能互开；在其它情形下，不可能互开。

原来，销售部门在一批锁具中随意地取每60个装一箱出售。团体顾客往往买几箱到几十箱，他们抱怨购得的锁具会出现互开的情形。现聘你为顾问，回答并解决以下的问题：

1) 每一批锁具有多少个，装多少箱。

2) 为销售部门提出一种方案，包括如何装箱(仍是60个锁具一箱)，如何给箱子以标志，出售时如何利用这些标志，使团体顾客不再或减少抱怨。

3) 采取你提出的方案，团体顾客的购买量不超过多少箱，就可以保证一定不会出现互开的情形。

4) 按照原来的装箱办法，如何定量地衡量团体顾客抱怨互开的程度(试对购买一、二箱者给出具体结果)。

1995年全国在学生数学建模竞赛

A题　一个飞行管理问题

在约10 000米高空的某边长160公里的正方形区域内，经常有若干架飞机作水平飞行。区域内每架飞机的位置和速度向量均由计算机记录其数据，以便进行飞行管理。当一架欲进入该区域的飞机到达区域边缘时，记录其数据后，要立即计算并判断是否会与区域内的飞机发生碰撞。如果会碰撞，则应计算如何调整各架(包括新进入的)飞机飞行的方向角，以避免碰撞。现假定条件如下：

1) 不碰撞的标准为任意两架飞机的距离大于8公里；

2) 飞机飞行方向角调整的幅度不应超过30度；

3) 所有飞机飞行速度均为每小时800公里；

4) 进入该区域的飞机在到达区域边缘时，与区域内飞机的距离应在60公里以上；

5) 最多需考虑6架飞机；

6) 不必考虑飞机离开此区域后的状况。

请你对这个避免碰撞的飞行管理问题建立数学模型，列出计算步骤，对以下数据进行计算(方向角误差不超过0.01度)，要求飞机飞行方向角调整的幅度尽量小。

设该区域4个顶点的座标为(0,0)，(160,0)，(160,160)，(0,160)。

记录数据为：

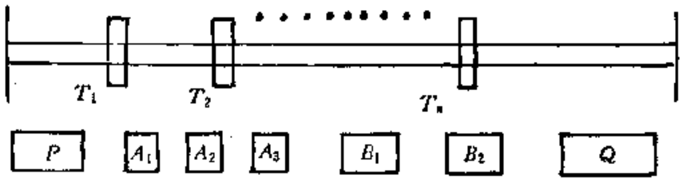
飞机编号	横座标 X	纵座标 Y	方向角(度)
1	150	140	243
2	85	85	236
3	150	155	220.5
4	145	50	159
5	130	150	230
新进入	0	0	52

注：方向角指飞行方向与 X 轴正向的夹角。

试根据实际应用背景对你的模型进行评价与推广。

B 题　天车与冶炼炉的作业调度

某钢铁厂冶炼车间的厂房布局是，地面沿一直线依次安置着 7 个工作点：辅料供应处 P；A 组 3 座转炉(冶炼成品钢)A_1,A_2,A_3；B 组 2 座冶炼炉(冶炼半成品钢，简称半钢)B_1，B_2；原料供应处 Q。这些设备的上方贯通着一条运送物料的天车轨道，上面布置着若干天车 $T_1,T_2,\cdots,T_n$ 为炉子作业服务，布局示意图如下。



天车与冶炼炉的作业过程与工序为：天车从 Q 处吊起原料一罐(吊罐时间 t_y)运至 B_1 或 B_2 处放下(放罐时间 t_i)，并将上一炉的原料空罐吊起(吊空罐时间 t_o)返回 Q 处放下(放空罐时间 t_k)。B 组炉在原料罐放下后即可在辅助作业下开始冶炼(冶炼时间 t_b)，冶炼后将半钢倒入空半钢罐(时间计入 t_b)，由天车吊起半钢罐(吊罐时间 t_d)运至 A_1 或 A_2、A_3 处将半钢倒入转炉(倒入时间 t_e)，并将空罐返回 $B1$ 或 $B2$ 处放下(放空罐时间 t_c)。再由天车从 P 处吊起辅料一槽(吊槽时间 t_g)运至 A_1 或 A_2、A_3 处加入转炉(加入时间 t_f)，并将空槽返回 P 处放下(放空槽时间 t_h)。A 组炉在半钢和辅料加入后即可开始冶炼(冶炼时间 t_a)，冶炼后成品钢的输出不用天车(输出时间计入 t_a)。天车通过相邻两个工作点的运行都相同，记为 t_x。

由于各台天车在同一条轨道上运行，因此其顺序位置 $T_1,T_2,\cdots,T_n$ 不可交换。在同一时间同一座炉子上只能允许一台天车作业；但 P、Q 两处可以允许多台天车同时作业。在 $P,A_1,\cdots,Q$ 每两个相邻工作点之间最多能容纳 2 台天车同时停放。

天车与冶炼炉作业调度的要求为：(1) 成品钢产量尽量高；(2) 各台天车的作业率(天车作业时间所占比例)尽量均衡(考虑到设备及人员安全等因素，一般天车作业率不超过 70%)；(3) 绝不允许出现天车相撞等事故；(4) 调度规定尽量简明，以利于现场人员使用。

现设定：$t_a=48,t_b=27,t_i=3,t_o=2,t_c=2,t_d=3,t_e=5,t_f=2,t_g=2,t_h=1,t_y=3,t_k=2$(单位：分钟)，$t_x=15$ 秒；A 组炉平均每炉产量 $W_a=120$ 吨。在不超过 5 台天

车的条件下，设计一种满足上述要求的天车与冶炼炉的作业调度方案：

(1) 各台天车负责哪些作业(列出《工序清单》)；

(2) 在所给方案的一个运行周期内，每一时刻天车和冶炼炉处于什么状态(画出《天车—炉子作用运行图》)；

(3) 一份供现场人员使用的《调度规则说明书》；

(4) 在所给方案下计算各台天车的作业率。

并按每天冶炼炉数估计该车间成品钢的年产量(扣除设备维修日，每台转炉作业日每年按300天计算)。

实际生产过程中，$t_a,t_b,\cdots,t_k$都是随机的(上面设定的数值可视为平均值)，讨论你的调度方案如何适用于实际生产过程。试提出该车间提高钢产量到年产300万吨以上的建议。

1996年全国在学生数学建模竞赛

A题　最优捕鱼策略

为了保护人类赖以生存的自然环境，可再生资源(如渔业、林业资源)的开发必须适度。一种合理、简化的策略是，在实现可持续收获的前提下，追求最大产量或最佳效益。

考虑对某种鱼(鳀鱼)的最优捕捞策略：

假设这种鱼分4个年龄组：称1龄鱼，……，4龄鱼。各年龄组每条鱼的平均重量分别为5.07，11.55，17.86，22.99(克)；各年龄组鱼的自然死亡率均为0.8(1/年)；这种鱼为季节性集中产卵繁殖，平均每条4龄鱼的产卵量为1.109×10^5(个)，3龄鱼的产卵量为这个数的一半，2龄鱼和1龄鱼不产卵，产卵和孵化期为每年的最后4个月；卵孵化并成活为1龄鱼，成活率(1龄鱼条数与产卵总量n之比)为$1.22\times10^{11}/(1.22\times10^{11}+n)$。

渔业管理部门规定，每年只允许在产卵孵化期前8个月内进行捕捞作业。如果每年投入的捕捞能力(如渔船数、下网次数等)固定不变，这时单位时间捕捞量将与各年龄组鱼群条数成正比，比例系数不妨称捕捞强度系数。通常使用13mm网眼的拉网，这种网只能捕劳3龄鱼和4龄鱼，其两个捕捞强度系数之比为0.42：1。渔业上称这种方式为固定努力量捕捞。

1) 建立数学模型分析如何可持续捕获(即每年开始捕捞时渔场中各年龄组鱼群条数不变)，并且在此前提下得到最高的年收获量(捕捞总重量)。

2) 某渔业公司承包这种鱼的捕捞业务5年，合同要求5年后鱼群的生产能力不能受到太大破坏。已知承包时各年龄组鱼群数量分别为：122，29.7，10.1，3.29($\times10^9$条)，如果仍用固定努力量的捕捞方式，该公司应采怎样的策略才能使总收获量最高。

B题　节水洗衣机

我国淡水资源有限，节约用水人人有责。洗衣在家庭用水中占有相当大的份额，目前洗衣机已非常普及，节约洗衣机用水十分重要。假设在放入衣物和洗涤剂后洗衣机的运行过程为：加水—漂洗—脱水—加水—漂洗—脱水—……—加水—漂洗—脱水(称

"加水 — 漂洗 — 脱水"为运行一轮"。请为洗衣机设计一种总用水量最少。选用合理的数据进行计算。对照目前常用的洗衣机的运行情况，对你的模型和结果作出评价。

1997年全国大学生数学建模竞赛题目

A题 零件的参数设计

一件产品由若干零件组装而成，标志产品性能的某个参数取决于这些零件的参数。零件参数包括标定值和容差两部分。进行成批生产时，标定植表示一批零件该参数的平均值，容差则给出了参数偏离其标定值的容许范围。若将零件参数视为随机变量，则标定值代表期望值，在生产部门无特殊要求时，容差通常规定为均方差的3倍。

进行零件参数设计，就是要确定其标定值和容差。这时要考虑两方面因素：

一、当各零件组装成产品时，如果产品参数偏离预先设定的目标值，就会造成质量损失，偏离越大，损失越大。

二、零件容差的大小决定了其制造成本，容差设计得越小，成本越高。

试通过如下的具体问题给出一般的零件参数设计方法。

粒子分离器某参数(记作 y)由7个零件的参数(记作 $x_1, x_2, \cdots, x_7$)决定，经验公式为

$$y = 174.42\left(\frac{x_1}{x_5}\right)\left(\frac{x_3}{x_2 - x_1}\right)^{0.85} \times \sqrt{\frac{1 - 2.62\left[1 - 0.36\left(\frac{x_4}{x_2}\right)^{-0.56}\right]^{3/2}\left(\frac{x_4}{x_2}\right)^{1.16}}{x_6 x_7}}$$

y 的目标值(记作 y_0)为1.50。当 y 偏离 $y_0 \pm 0.1$ 时，产品为次品，质量损失为1 000(元)；当 y 偏离 $y_0 \pm 0.3$ 时，产品为废品，损失为9 000(元)。

零件参数的标定值有一定的容许变化范围；容差分为 A、B、C 三个等级，用与标定值的相对值表示，A 等为 $\pm 1\%$，B 等为 $\pm 5\%$，C 等为 $\pm 10\%$ 7个零件参数标定值的容许范围，及不同容差等级零件的成本(元)如下表(符号 / 表示无此等级零件)：

	标写值容许范围	C 等	B 等	A 等
x_1	[0.075, 0.125]	/	25	/
x_2	[0.225, 0.375]	20	50	/
x_3	[0.075, 0.125]	20	50	200
x_4	[0.075, 0.125]	50	100	500
x_5	[1.125, 1.875]	50	/	/
x_6	[12, 20]	10	25	100
x_7	[0.562 5, 0.935]	/	25	100

现进行成批生产，每批产量1 000个，在原设计中，7个零件参数的标定值为：$x_1 = 0.1, x_2 = 0.3, x_3 = 0.1, x_4 = 0.1, x_5 = 1.5, x_6 = 16, x_7 = 0.75$；容差均取最便宜的等级。

请你综合考虑 y 偏差 y_0 造成的损失和零件成本，重新设计零件参数(包括标定值和容差)，并与原设计比较，总费用降低了多少。

B题　截断切割

某些工业部门(如贵重石材加工等)采用截断切割的加工方式。这里"截断切割"是指将物体沿某个切割平面分成两部分。从一个长方体中加工出一个已知尺寸、位置预定的长方体(这两个长方体的对应表面是平行的),通常要经过6次截断切割。

设水平切割单位面积的费用是垂直切割单位面积费用的 r 倍,且当先后两次垂直切割的平面(不管它们之间是否穿插水平切割)不平行时,因调整刀具需额外费用 e。

试为这些部门设计一种安排各面加工次序(称"切割方式")的方法,使加工费用最少。(由工艺要求,与水平工作台接触的长方体底面是事先指定的)详细要求如下:

1)需考虑的不同切割方式的总数。

2)给出上述问题的数学模型和求解方法。

3)试对某部门用的如下准则作出评价:每次选择一个加工费用最少的待切割面进行切割。

4)对于 $e=0$ 的情形有无简明的优化准则。

5)用以下实例验证你的方法:待加工长方体和成品长方体的长、宽、高分别为10、14.5、19和3、2、4,二者左侧面、正面、底面之间的距离分别为6、7、9(单位均为厘米。)。垂直切割费用为每平方厘米1元,r 和 e 的数据有以下4组:

$$a.\ r=1, e=0;\qquad b.\ r=1.5, e=0;$$
$$c.\ r=8, e=0;\qquad d.\ r=1.5; 2\le e\le 15。$$

对最后一组数据应给出所有最优解,并进行讨论。

参考文献

[1] 姜启源. 数学模型. 北京:高等教育出版社,1987

[2] 美 JWILLIAM F Lucas. 微分分程模型. 朱煜民,周宇虹译. 国防科技大学出版社,1988

[3] 陈兰荪. 数学生态学模型与研究方法. 科学出版社,1998

[4] 沈继红等. 数学建模. 哈尔滨工程大学出版社,1996

[5] 谌安琦. 科技工程中的数学模型. 中国铁道出版社,1998

[6] 李庆杨、王能超、易大义编. 数值分析. 华中理工大学出版社,1982

[7] 曹瓔珞、曹德欣编. 计算方法. 中国矿业大学出版社,1994

[8] 孙家昶著. 样条函数与计算几何. 科学出版社,1982

[9]《运筹学》教材编写组. 运筹学. 清华大学出版社,1982

[10] 范鸣玉、张莹编著. 最优化技术基础. 清华大学出版社,1982.11

[11] 王保伦、石维明、革永才等编著. 矿业实用运筹学. 东北工学院出版社,1991.10

[12] 南京地区工科院校数学建模与工业数学讨论班编. 数学建模与实验. 河海大学出版社. 1996.11

[13] 宋念慈编著. 工商管理数理统计. 上海交通大学出版社. 1994.4

[14] 寿纪麟. 数学建模—方法与范例. 西安交通大学出版社. 1993.12

[15] 孙东升等编著. 系统工程简明教程. 湖南科学技术出版社. 1987.5

[16] 禹小平等. 逐步回归在牲猪生产决策中的应用. 数理统计与管理. 1992.5

[17] 赵怀章等编著. 实用市场经济数学模型. 同济大学出版社. 1997.5

[18] 卢开澄等编著. 图论及其应用. 清华大学出版社. 95.8

[19] 肖位枢主编. 图论及其算法. 航空工业出版社. 93.7